

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 26, 2002, 10:19:57 ; Search time 3268 Seconds
(without alignments)
4247.870 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPFRAGVNVGVDM.....FEDGCTWNIKKIQDQKKG 477

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USFTO.spool/US0912020/runat_25112002_091428_24247/app_query.fasta_1.647
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0912020.ecgn_1.1.3637 @runat_25112002_091428_24247 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2393	100.0	1434	6	AX189075	Sequence
2	2393	100.0	1434	6	AX363670	Sequence
3	2393	100.0	11307	1	AE000387	Escherich
4	2393	100.0	25638	1	ECU28379	Escherich
5	2385	99.7	12088	1	AE005534	Escherich
6	2385	99.7	307962	1	AP002564	Escherich
7	2276	95.1	1431	1	AF155126	Salmonell
8	2276	94.7	20604	1	AE008847	Salmonell
9	2265	94.7	258050	1	AL627278	Salmonell
10	2250	94.0	1476	1	AF163661	Salmonell
11	2225	93.0	1464	1	AF163662	Salmonell
12	2076	86.8	10641	1	AE013955	Yersinia
13	2076	86.8	208050	1	AJ414144	Yersinia
14	1712	71.5	14465	1	U32828	Haemophilus
15	1701	71.1	10747	1	AE006127	Pasteurel
16	1637	68.4	10373	1	AE004313	Vibrio ch
17	1466	61.3	2431	1	HU17642	Haemophilus
18	1350.5	56.4	12498	1	AE004912	Pseudomon
19	1036	43.3	1001	6	AX081559	Sequence
20	1036	43.3	1001	6	AX374740	Sequence
21	904.5	37.8	10637	1	AE006022	Caulobact
22	863	36.1	282183	1	CJ11168X4	Campyloba
23	863	36.1	348077	1	AP003000	Mesorhizo
24	819.5	34.2	19549	1	AE001509	Helicobac
25	811	33.9	347550	1	AP001118	Buchnera
26	809.5	33.8	23147	1	AE000596	Helicobac
27	806	33.7	13965	1	AE014081	Buchnera
28	753.5	31.5	3780	1	AF125564	Neisseria
29	753.5	31.5	11133	1	AE002435	Neisseria
30	753.5	31.5	349980	6	AX044030	Sequence
31	745.5	31.2	969	6	AX300225	Sequence
32	745.5	31.2	4409	1	NNE391258	Neisseria
33	745.5	31.2	311321	1	NMA332491	Neisseria
34	721.5	30.2	198050	1	AL646061	Ralstonia
35	680.5	28.4	3530	1	AF159428	Burkholde
36	610.5	25.5	77457	1	AF210249	Streptomy
37	594	24.8	38404	1	SC2G5	Streptomy
38	584.5	24.4	10107	1	AE010481	Fusobacte
39	566	23.7	13591	1	AE000696	Aquifex a
40	492.5	20.6	11460	1	AE012878	Chlorobiu
41	458	19.1	298750	1	AP005375	Thermosyn
42	413.5	17.3	12113	1	AE010600	Fusobacte
43	411.5	17.2	347	6	AR203746	Sequence
44	396	16.5	10548	1	AE012800	Chlorobiu
45	391.5	16.4	15451	1	AE000677	Aquifex a

ALIGNMENTS

RESULT 1

AX189075
LOCUS AX189075 1434 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 276 from Patent WO0148209.
ACCESSION AX189075
VERSION AX189075.1 GI:15142798
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 1434)
REFERENCE Forsyth, R.A., Ohlsen, K.L. and Zyskind, J.W.
AUTHORS Genes identified as required for proliferation of E. coli
TITLE Patent: WO 0148209-A 276 05-JUL-2001;
JOURNAL Elitra Pharmaceuticals, Inc. (US)
FEATURES
source location/Qualifiers
1..1434
/organism="Escherichia coli"
/db_xref="taxon:562"
1..1434
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC50756.1"
/db_xref="GI:15142798"
translation="MKVTLPEFERAGVMVGDVMDLRYWGYGPTSRISPEAPVPVVKVN
TIERPGGAANVAMNISLANRVLGLTGDAAARLSKSLADVNVKDFVSVPTHP
TITKLVLSRNOQLRLDFEGFEVGPOLPLHERINOALSISGIALVLSYAKGALASV
COMIOLARKAGPVLLIDPKGTDFFERYGATLLTPNLSFEAVVCKKTEELIVERGMK
LIADYELSAVVTSRQGMSSLLQPKAPLHPTQAOEYVDVTGAGDTVIGVLAATLAA
GNSLEAEAFANAAGVVGKLGSTVSPFIELENAVVRADTGFQVMTTEELKLVAA
ARKREKVVYNGVFDILHAGVSYLANARKLDRLIVAVNSDASTKRLKGDSPVNP
LEQRMIVLEADVWYVSFEEDTPQRLIAGTLPDLLVKGDDYKPEETAGSKVWANG
GEVLNFDGSGSTNIIKKIQDQKKG"
BASE COUNT 328 a 352 c 447 g 307 t
ORIGIN
Alignment Scores:
Pred. No.: 1.27e-138 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-912-020-325 (1-477) x AX189075 (1-1434)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
DB 1 ATGAAAGTAACGCTGCAGAGTTTGAACGTGCAGGAGTGTGGTGGTGTGTGTGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 61 CTGGATCGTTACTGTGACGCCGCCACCACTGTATCTCCCGGAGCGCGGTGCGCGT 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
DB 121 GTTAAAGTGAATACCATCAAGAACCTCCGGCGCGCGCTAACCTGGCGATGAATATC 180
QY 61 AlaSerLeuGluGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
DB 181 GCTTCCTCGGTGCTAATGCAGCCCTGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGT 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAlaAsnValLysCysAspPheValSerValProThr 100
DB 241 GCCTGAGTAATCTCTGGCGGACAGTCAACGTCAAATGCGACTTCGTTCTGTACCGACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
DB 301 CATCCGACCATTAACAAATACGGGTACTTCCCGCAACCAACAGCTGATCCGCTCTGGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
|||||

Db 361 TTTGAAGAAGGTTTCAAGGTTGTGATCCGCGACCGCTGCACGAGCGGATTAATCAGCGC 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
|||||
Db 421 CTGAGTTGAGTGGCGCGCTGCTTCTGACTACGCCAAAGGTCGCTGCGCAAGCGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
|||||
Db 481 CAGCAGATGATCAACTGCGCGTAAAGCGGTGTTCGCGTCTGATTGATCAAAAGGT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
|||||
Db 541 ACCGATTTTGAAGCGTACCGCGCGCTACGCTGTAAACCGCGAATCTCTCGGAATTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
|||||
Db 601 GCTGTGTCGGTAAATGTAAAGACCGAAGAGAGATTGTTGACGCGCGCATGAACATCAT 560
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlnMetSerLeuLeu 240
|||||
Db 661 GCCGATTACGAACTCTCGGCTCTGTAGTGACCCGTTCCGAACAGGGTATGTGCGTCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
|||||
Db 721 CAACCGGGTAAAGCGCGCTGCTATATGCCAACCCAAAGCGCAGAGTGTATGACGTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
|||||
Db 781 GGTGCGGCGCACACGCTGATTGGCGTCTGCGGCGCAACGCTGGCAGCGGTAAATCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
|||||
Db 841 GAAGAAGCCTGCTCTTCTCCCAATGCGCGCGTGGCGTGGTGGTGGTGGTGGTGGTGG 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
|||||
Db 901 TCCACGGTTTCGCGCATCGAGCTGGAAATGCTGTACGTGGAGCTGCATACAGGCTTT 960
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
|||||
Db 961 GCGGTGATCACCGAAGAGAACTGAAGCTGGCGTAGCGGACGCGGCTAAACGTGGTGAA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
|||||
Db 1021 AAAGTGGTGATGACCAACGCTGCTTTGTACATCTCTGACGCGCGGCGACGCTCTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAlaAsnSerAspAlaSerThr 380
|||||
Db 1081 GCAAATGCCCGCAAGCTGGTGACCGCTTGTATTGTTGCCGTCAACAGCGATGCTCCACCC 1140
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
|||||
Db 1141 AAAGGCTCAAAAGGGGATTCCCGCGGTAAACCCACTCGAACAGCTATGATTGTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
|||||
Db 1201 GGCACACTCGAAGCGGTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1260
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
|||||
Db 1261 ATCGCGCGGATCTGCGCAGATCTGCTGGTGGTAAAGCGCGGACTATTAACAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
|||||
Db 1321 GCCGGGAGTAAAGAGTCTGGGCCAACGCTGGCGAAGTGTGTGCTCAACTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysGly 477
|||||
Db 1381 GGTGTGTCGACGACCAACATCATCAAGAAGATCCAAACAGGATAAAAAAGGC 1431
RESULT 2
AX363670 1434 bp DNA linear PAT 15-FEB-2002
LOCUS

DEFINITION Sequence 165 from Patent EP1178052.
 ACCESSION AX363670
 VERSION AX363670.1 GI:18695784
 KEYWORDS Escherichia coli.
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

FEATURES
 Location/Qualifiers
 1..1434
 /organism="Escherichia coli"
 /db_xref="taxon:562" 307 t

BASE COUNT 328 a 352 c 447 g 307 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,27e-138 Length: 1434
 Score: 2393.00 Matches: 477
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-912-020-325 (1-477) x AX363670 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
 |||||
 Db 1 ATGAAGTAAACGTCGCAGAGTTGAACGTCAGAGGAGTGATGGTGTGTGTGTATG 60
 |||||

QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
 |||||
 Db 61 CTGGATCGTTACTGGTACGCGCCACCACGATCGTATCTCGCCGGAAGCGCGTGGCGTG 120
 |||||

QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
 |||||
 Db 121 GTTAAAGTGAATACCATCGAAGAACGTCGCGGCGCGGCTAACGTTGCGGATGATATC 180
 |||||

QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
 |||||
 Db 181 GCTTCTCGGTCTAATGACCGCTGCTGGTGTGACGGGCATGTACGATGACGCGCGC 240
 |||||

QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
 |||||
 Db 241 GCCTGAGTAAATCTCTGCGCGACGTCACGTCACAAATGCGACTTCGTTTCTGTACCGACG 300
 |||||

QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
 |||||
 Db 301 CATCCGACCATTAACAAATACCGGGTACTTTCGCCAACCAACAGCTGATCCGCTCTGGAT 360
 |||||

QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
 |||||
 Db 361 TTTGAGAAGTTTCGAGAGTGTGATCCGACCGCGCTGCAGCGGATTAATACGCGC 420
 |||||

QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
 |||||
 Db 421 CTGAGTTGATTTGGCGCGTGGTGTCTTCTGACTACGCCAAAGGTGGCGTGGCAAGGTA 480
 |||||

QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
 |||||
 Db 481 CAGCAGATGATCCAACTGGCGCGCTAAAGCGGGTGTTCGCGTGTGATTGATTCGCAAAAGGT 540
 |||||

QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
 |||||
 Db 541 ACCGATTTGAGCGCTACCGCGCGCTACCGTGTAAACCGCGGATCTCTCGGAATTTGAA 600
 |||||

QY 201 AlaValValGlyLysCysLysThrGluGluLeuIleValGluArgGlyMetLysLeuIle 220
 |||||
 Db 601 GCTGTTGTCGTAATGTAAGACCGAAGAGAGATTGTTGAGCGCGCATGAACTGATT 660
 |||||

QY 221 AlaAspTyrGlnLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
 |||||
 Db 661 GCGGATTACGAACCTCTCGGCTCTGTTAGTAGCCCGTTCGGAACAGGGGTATGCTGCTGCTG 720
 |||||

QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
 |||||
 Db 721 CAACCGGTAAGCGCGCTGCATATGCCAACCAACCGCAGGAAGTGTATGACGTTACC 780
 |||||

QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
 |||||
 Db 781 GGTGCGCGCACACGGTGTATGGCTCTGCGGCAACGCTGGCAGCGGGTAATTCGCTG 840
 |||||

QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
 |||||
 Db 841 GAAGAAGCCTGCTCTTTTGCCAATGCGCGCTGGCTGGTGGTGGCGCAACTGGGAACC 900
 |||||

QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
 |||||
 Db 901 TCCACGGTTTCGCGATCGAGCTGGAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
 |||||

QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaAlaArgLysArgGlyGlu 340
 |||||
 Db 961 GCGGTGATGACCGAAGAGGAACCTGAAGCTGCGCTAGCGCGCAGCGCTAAACGTTGGTAA 1020
 |||||

QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
 |||||
 Db 1021 AAAAGTGTGATGACCAACGGTGTCTTTGACATCTCGACCGCGGCGACGCTCTTTATCTG 1080
 |||||

QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaAlaValAsnSerAspAlaSerThr 380
 |||||
 Db 1081 GCRAATCCCGCAAGCTGGGTGACCGCTTGTATTGTCGCGTCAACACGATGATCCACCC 1140
 |||||

QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
 |||||
 Db 1141 AAACGGCTGAAGGGGATTCGCGCGGTAAACCCACTCGAACACGCGTATGATTGTGCTG 1200
 |||||

QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
 |||||
 Db 1201 GCGCACCTGAACGGTTCGACTGGGTAGTGTCTGTTTGAAGAGGACACGCGCGACGCTG 1260
 |||||

QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
 |||||
 Db 1261 ATCGCGGGATCTTGCAGATCTGCTGGTGAAGCGCGCGACATATAACCAAGAGAGATT 1320
 |||||

QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 |||||
 Db 1321 GCGGGAGTAGTAAGAAGTCTGGCCACGCTGGCGAAGTGTGTGCTCAACTTTTGAAGAC 1380
 |||||

QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnAspLysLysGly 477
 |||||
 Db 1381 GGTGCTCGACGACCAACATCATCAAGAAGATCCACAGGATAAAAAAGGC 1431
 |||||

RESULT 3
 AE000387/c 11307 bp DNA linear BCT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 277 of 400 of the complete genome.
 DEFINITION
 ACCESSION AE000387 U00096
 VERSION AE000387.1 GI:1789431
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 11307)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Bernal, N.T., Burland, V., Riley, N., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11307)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,

University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecologi@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

3 (bases 1 to 11307)
 Direct Submission
 Blattner, F.R.
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecologi@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

4 (bases 1 to 11307)
 Plunkett, G. III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MGI655. Predicted open reading
 frames were determined using GeneMark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 (e-mail: mark@camber.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES

source
 1. .11307
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MGI655"
 /db_xref="taxon:83333"
 complement(240. .1673)
 /gene="b3052"
 complement(240. .1673)
 /gene="b3052"
 /function="putative enzyme: Not classified"
 /note="f477"
 /codon_start=1
 /transl_table=11
 /product="putative kinase"
 /protein_id="AAC76088.1"
 /db_xref="GI:1789432"
 /translation="MKVTLFEFFRAGVWVGVDMVDRLYWGTPSTRISPEAPVPVVKYN
 TIERPGGANVANNIASLGRNVLGLTGIDDAARALSKSLADVNVKDFVSPVTPH
 TTKRLVSRNQILRLDFEGFEGVDPQLHERINQALSSIGALVSLDYGAKGALASV
 QITQLARKAGVPVLPIDPKGTDYRGATLLTPNLSEFAVVGKCKTEIEIYERGMK
 LIADYELALLVTESEGMSILOPGKAPLHMPYQAOEYVDVTGAGDVIQVLAATLAA
 NSLEECACFPANAGVGVVKGSTVSPTELENAVGRADTGGVMTREELKLAIVAA
 ARKGRKVTNGVGLLHAGKSYLANARKGLDRLLIVAVNSDASRKLKGDSPVNP
 LEQRMIVIGALEAVDGVVSEEDTPQRLIAGILPDLVLKVGDKYKPEIAGSKYEWANG
 GEVLVLFNFGDCSTNIIKKIQDQKG"
 complement(1695. .1723)
 /note="factor Sigma70; predicted +1 start at 3194409"
 complement(1721. .4561)
 /gene="glnE"
 /note="b3053"

promoter

gene

CDS

complement(1721. .4561)
 /gene="glnE"
 /function="enzyme: Proteins - translation and
 modification"
 /note="f946; 99 pct identical (1 gap) to GLNE_ECOLI
 SW:P30870"
 /codon_start=1
 /transl_table=11
 /product="adenylating enzyme for glutamine synthetase"
 /protein_id="AAC76089.1"
 /db_xref="GI:1789433"
 /translation="MKPLSSPQQYWTQVTLVVERLPEPLAEESLSQAQSVLTFSDVQD
 VAIAPHLWTELESQPPDAEWHYAALQALCNVSEAGLWELRFLRRIRWVRIA
 WAOTLAIVTESILQQLSYLAETLIVAARDLWYDACCREWCTPCNAQCEAPQLILGM
 KGLGELNFSSDILLFANPEHCGTGGRELDNAQGFTRMGORLILVLPQDQGF
 VYVDMRLRPFGEESPLVFAALEDIYQEGGRWERYAVYKARIMGSEIYVQVNFQ
 AMLRPFVFRYIDFSVIQSLRNKMGMIAREVRRGLTDNIKLAGGIRSEIYFVQVFO
 LIRGPEPSQSRLPTLSAIELHLLSENDAEQRYAYFLRRLLENLQGSINDEQT
 QTLPSDELNARLAWADFADWPQTLTGAHTMTNRVYFNLIGDDESEQESLSSE
 QWRLEWDALQEDDTVPVLAHLSDDRRQVLTLDADFKELDKRTIGRGGYDQHLH
 PHLLSCAREDAATVTSRILALVIVTRTTLLELSEFPAAKLHLISLCAASPMIA
 SOLARYPLLDLDELDPNTLYOPTATDAYRDELQYLLRVLPEDDEQLEALRQPKQAQ
 LLRTAAADIAGTLPVMKYSDHLTLWLAEMIDAVVQQAQVQVARIKPNHLEREGRG
 FAVVYKGLGWEIGYSDUDLILFHDCPMDAMTDGERIDGROFYLRQAQRIMHLFS
 TRTSSGLIYEDARLPRGAGMLVTSAEAFADYQKNEAWTWEHALVRAARVYGDQ
 LTAHFDAVRREIMTLPRGKTLQTEVREMRKRAHLNCKHRDFDIKADGGITDIE
 FITOYLVRVAHEKPKLTRNSDNRVILLELAQNDIMEQEAAMALTRAYTTLRDELHLH
 ALQPLGHVSDCTAERLVRASQWKLVEE"
 complement(4584. .5885)
 /gene="ygiP"
 /note="b3054"
 complement(4584. .5885)
 /gene="ygiP"
 /function="orf; Unknown"
 /note="f433"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC76090.1"
 /db_xref="GI:1789434"
 /translation="MAQIEIKLFIVNHSVAELRDHLNLTGGEHHPVOLLNIYETP
 DNWLGRHDMGLRIRGREGVMTKVGAVRTGGLHQRYEYNVALSEPTLDIAQLPTEV
 WPNGLPADLASRVQPLFTDFYREKWLVAVDGSOIETALDQGEVKAQFEAPICELE
 LELSGDTRAVLKLANQVOTLROGSLSKAARGYHLAQNAPAREIKPTTILHVAAK
 ADVEQGLEALELAAQWYHEELVRSNDAKAEQVLAALSLVRHTLMFGIYVPRKA
 STHLRDLTQCEATIASVNAVYATYSTETAMAKALITELWLYSKAWQFPDLDAKQGGKI
 SDSKRFADILHSRHAELKSVFCOPLGDRYRDQPLRTLDIDSLILLAGYDYPVVAQ
 AWLENWQGLHAIATGQRIEIEHFRNEANNOEPFWLHSGKR"
 complement(5878. .5902)
 /note="No predicted promoter"
 /bound_moiety="Cynr predicted site"
 5932. .5954
 /note="central position to predicted promoter: -154.5"
 /bound_moiety="Fur predicted site"
 5932. .5954
 /note="central position to predicted promoter: -38.5"
 /bound_moiety="Fur predicted site"
 5947. .5976
 /note="factor Sigma70; predicted +1 start at 3198704"
 6064. .6092
 /note="factor Sigma70; predicted +1 start at 3198820"
 6122. .6155
 /note="REP (repetitive extragenic palindromic) element;
 contains 1 REP sequence"
 6127. .6747
 /gene="ygiM"
 /note="b3055"
 6127. .6747
 /gene="ygiM"
 /function="orf; Unknown"
 /note="o206; 100 pct identical to 127 amino acids of
 YGIM_ECOLI SW: P39202 but has 79 additional N-ter

CDS

complement(1721. .4561)
 /gene="glnE"
 /EC_number="2.7.7.42"
 /function="enzyme: Proteins - translation and
 modification"
 /note="f946; 99 pct identical (1 gap) to GLNE_ECOLI
 SW:P30870"
 /codon_start=1
 /transl_table=11
 /product="adenylating enzyme for glutamine synthetase"
 /protein_id="AAC76089.1"
 /db_xref="GI:1789433"
 /translation="MKPLSSPQQYWTQVTLVVERLPEPLAEESLSQAQSVLTFSDVQD
 VAIAPHLWTELESQPPDAEWHYAALQALCNVSEAGLWELRFLRRIRWVRIA
 WAOTLAIVTESILQQLSYLAETLIVAARDLWYDACCREWCTPCNAQCEAPQLILGM
 KGLGELNFSSDILLFANPEHCGTGGRELDNAQGFTRMGORLILVLPQDQGF
 VYVDMRLRPFGEESPLVFAALEDIYQEGGRWERYAVYKARIMGSEIYVQVNFQ
 AMLRPFVFRYIDFSVIQSLRNKMGMIAREVRRGLTDNIKLAGGIRSEIYFVQVFO
 LIRGPEPSQSRLPTLSAIELHLLSENDAEQRYAYFLRRLLENLQGSINDEQT
 QTLPSDELNARLAWADFADWPQTLTGAHTMTNRVYFNLIGDDESEQESLSSE
 QWRLEWDALQEDDTVPVLAHLSDDRRQVLTLDADFKELDKRTIGRGGYDQHLH
 PHLLSCAREDAATVTSRILALVIVTRTTLLELSEFPAAKLHLISLCAASPMIA
 SOLARYPLLDLDELDPNTLYOPTATDAYRDELQYLLRVLPEDDEQLEALRQPKQAQ
 LLRTAAADIAGTLPVMKYSDHLTLWLAEMIDAVVQQAQVQVARIKPNHLEREGRG
 FAVVYKGLGWEIGYSDUDLILFHDCPMDAMTDGERIDGROFYLRQAQRIMHLFS
 TRTSSGLIYEDARLPRGAGMLVTSAEAFADYQKNEAWTWEHALVRAARVYGDQ
 LTAHFDAVRREIMTLPRGKTLQTEVREMRKRAHLNCKHRDFDIKADGGITDIE
 FITOYLVRVAHEKPKLTRNSDNRVILLELAQNDIMEQEAAMALTRAYTTLRDELHLH
 ALQPLGHVSDCTAERLVRASQWKLVEE"
 complement(4584. .5885)
 /gene="ygiP"
 /note="b3054"
 complement(4584. .5885)
 /gene="ygiP"
 /function="orf; Unknown"
 /note="f433"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC76090.1"
 /db_xref="GI:1789434"
 /translation="MAQIEIKLFIVNHSVAELRDHLNLTGGEHHPVOLLNIYETP
 DNWLGRHDMGLRIRGREGVMTKVGAVRTGGLHQRYEYNVALSEPTLDIAQLPTEV
 WPNGLPADLASRVQPLFTDFYREKWLVAVDGSOIETALDQGEVKAQFEAPICELE
 LELSGDTRAVLKLANQVOTLROGSLSKAARGYHLAQNAPAREIKPTTILHVAAK
 ADVEQGLEALELAAQWYHEELVRSNDAKAEQVLAALSLVRHTLMFGIYVPRKA
 STHLRDLTQCEATIASVNAVYATYSTETAMAKALITELWLYSKAWQFPDLDAKQGGKI
 SDSKRFADILHSRHAELKSVFCOPLGDRYRDQPLRTLDIDSLILLAGYDYPVVAQ
 AWLENWQGLHAIATGQRIEIEHFRNEANNOEPFWLHSGKR"
 complement(5878. .5902)
 /note="No predicted promoter"
 /bound_moiety="Cynr predicted site"
 5932. .5954
 /note="central position to predicted promoter: -154.5"
 /bound_moiety="Fur predicted site"
 5932. .5954
 /note="central position to predicted promoter: -38.5"
 /bound_moiety="Fur predicted site"
 5947. .5976
 /note="factor Sigma70; predicted +1 start at 3198704"
 6064. .6092
 /note="factor Sigma70; predicted +1 start at 3198820"
 6122. .6155
 /note="REP (repetitive extragenic palindromic) element;
 contains 1 REP sequence"
 6127. .6747
 /gene="ygiM"
 /note="b3055"
 6127. .6747
 /gene="ygiM"
 /function="orf; Unknown"
 /note="o206; 100 pct identical to 127 amino acids of
 YGIM_ECOLI SW: P39202 but has 79 additional N-ter

protein_bind

protein_bind

protein_bind

promoter

promoter

repeat_region

gene

CDS


```
residues"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76091.1"
/db_xref="GI:1789435"
/translation="WPKLRLIGLTLALSATVSHAEETRYVSDLENTWVRSGPGDHY
RLVGVNAGEEVTLLQTDANTNYAOVKDSSGTAMPLKQLSTEPSLSRVPDLNQV
KTLTKLNLIDNTNQRPAEQKVAQSDSVINGLKEENOKLKNELIVAQKVDASV
QDDKQRTIIMQWFMVGGVGLGILLGLVLPLPSRRKRDWMN"
6747..6775
/promoter
/feature="factor Sigma70; predicted +1 start at 3199503"
6811..8049
/gene
/gene="cca"
/feature="p3056"
6811..8049
/CDS
/gene="cca"
/feature="p3056"
/EC_number="2.7.25"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA
modification"
/feature="p412; 100 pct identical to CCA_ECOLI SW: P06961; CG
Site No. 933"
/codon_start=1
/transl_table=11
/product="tRNA nucleotidyl transferase"
/protein_id="AAC76092.1"
/db_xref="GI:1789436"
/translation="WKIYLGGAVRDALLGLPVKDRDWWVVGSTPOEMLDAGYQQVGR
DFPVLPHQTHEEYALARKTERKSGSYGFTCYAAPDVLTEDDLKRRDLTINALAODD
NGEILDYNGTDLQNLRLRHVSFAPGEDPLRVLRVAFARAYAHGLFRADIETLALM
REMHAGELEHLTPERVWKETESALTTRNPQVFFQVLRDCGALRVLFPEIDALFGVPA
```

Alignment Scores:

Pred. No.:	1,54e-137	Length:	11307
Score:	2393.00	Matches:	477
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-912-020-325 (1-477) x AE000387 (1-11307)

QY	1	MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet	20
Db	1673	ATCAAGATACGCTGCACAGTTTGAACGTGCAGGAGTGATGCTGCTGATGATG	1614
QY	21	LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProVal	40
Db	1613	CTGGATCGTTACTGTCACGCCCCACAGTCGTATCTCGCGGAAGCGCGTGC	1554
QY	41	ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle	60
Db	1553	GTTAAAGTGNATACCATCAGAACACGTCGGCGCGCGCGGTAACTGATGATATC	1494
QY	61	AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArg	80
Db	1493	GCTTCTCTCGTCTAATGACCCCTGCTGCGGTGACGGCATTCACGATGACGCGC	1434
QY	81	AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr	100
Db	1433	GCCTCAGTAAATCTCTGCGCGACGTCACGTCAAATGGCACTTCGTTCTGTACCGACG	1374
QY	101	HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp	120
Db	1373	CATCCGACCATPACCAATACCGGTACTTTCGCGCAACCAACAGCTGATCCGCTGGAT	1314
QY	121	PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla	140
Db	1313	TTTGAAGAAGGTTTGAAGGTGTGATCCGACGCCGTCGACGAGCGGATTAATCAGGCG	1254
QY	141	LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal	160
Db	1253	CTGAGTTCGATTGGCGCGTGGTCTTCTGACTACGCCAAAGGTGCGCTGGCAAGCGTA	1194

QY	161	GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly	180
Db	1193	CACGATGATCACTCACTGGCGGTAAAGCGGGTTCCTCCGGTGTGATTGATCAAAAAGGT	1134
QY	181	ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu	200
Db	1133	ACCGATTTCAGCGCTACCGCGGGCTACGCTCTTAACGCCGAATCTCTCGGAATTGAA	1074
QY	201	AlaValAlaGlyLysCysLysThrGluGluGluValGluArgGlyMetLysLeuIle	220
Db	1073	GCTGCTGCTGGTAAATCTAAGACCGAAGAGAGATTCTTCAGCGCGCATGAACACTGATT	1014
QY	221	AlaAspTyrClnLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeu	240
Db	1013	GCGGATTACAACTCTCGGCTCTGTAGTACCCCTTCCGACAGGATGTGCGCTGCTG	954
QY	241	GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr	260
Db	953	CAACCGGGTAAACCGCGCTGCATATGCCAACCCCAAGCGCAGCAAGTGTATGAGTTACC	894
QY	261	GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu	280
Db	893	GGTGGCGGCGACACGCTGATTGGCTCTCTGGCGCAACGCTGGCAGCGGTAATTCGCTG	834
QY	281	GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr	300
Db	833	GAAGAAGCTTGCTCTTTGCCAATCGCGCGCTGGCTGGCTGGCGCAACCTGGGAACC	774
QY	301	SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe	320
Db	773	TCCACGGTTTCGCGGATCGAGCTGGAAATGCTGTACGTGACGCTGCAGATACAGGCTTT	714
QY	321	GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu	340
Db	713	GGCGTGATCACCGAAGAGGAACCTGAAGCTGGCGGTAGCGCGCAGCGCTAAACGCTGGAA	654
QY	341	LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu	360
Db	653	AAAGTGGTATGACCAACCGGTGCTTTGACATCTCGACCCCGCGGACGCTCTCTATCTG	594
QY	361	AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr	380
Db	593	GCAATGCCGCGCAAGCTGGGTGACCGCTTCAATGTTGCCGTCAACACGATGCTCCACC	534
QY	381	LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu	400
Db	533	AAACGGCTGAAGGGGATTCGCCCGCGGTAAACCCATCTCGAACAGCGTATGATTGCTG	474
QY	401	GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu	420
Db	473	GGCGCATCGAAGCGGTGCGTGGTAGTGTCTGTTGAAGAGACACGCCGCGCGCTTG	414
QY	421	IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle	440
Db	413	ATCCCGGGATCTTTCGACATCTGCTGGTGAAGGGCGGCGACTATAAACCAAGAGAGATT	354
QY	441	AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp	460
Db	353	GCCGGGAGTAAAGAGTCTGGGCCAACGGTGGCGAAGTGTGGTGGCTCAACTTTGAAGAC	294
QY	461	GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLysGly	477
Db	293	GGTGTCTCGACGACCAACATCATCAAGAGATCCACAGGATTAATAAAGC	243

RESULT 4
ECU28379
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

ECU28379
Escherichia coli K-12 genome; approximately 68 minutes.
U28379.1
GI:1203797
Escherichia coli.

ORGANISM	Escherichia coli		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
AUTHORS	Escherichia.		
TITLE	1. (bases 1 to 25638)		
JOURNAL	Plunkett, G.		
COMMENT	Direct Submission Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA On Feb 27, 1996 this sequence version replaced gi:882576. This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HC00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. The end of this entry overlaps the start of the entry ECOUW67 (U18997) by 1596 bp. This entry should be considered somewhat provisional; it will be updated and merged with others at a later date.		
FEATURES	Location/Qualifiers		
source	1. .25638		
	/organism="Escherichia coli"		
	/strain="K-12"		
	/sub_strain="MG1655"		
	/db_xref="taxon:562"		
	/map="approximately 68 minutes"		
	/note="This sequence is from lambda clone		
	DD689 (EC27-1209); M13Janus vector was used for subcloning"		
CDS	6002..6622		
	/note="alternate name ygiM; ORF_0206"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89135.1"		
	/db_xref="GI:1203798"		
	/translation="MPKRLIGLTLALSAATVSHAEETRYVSDLELNTWVRSGPDHY RLVGTWAGSEVTLQLQDANTYAQAQKDSGRTAWIPKQLSTEPSLRKRPDLEQV KTYDKLTINDNTWORTAEQKVAQSDVINGLKEENQKLKLNELIYAKKKVDAASV QLDKQRTIIMQFMFYGGVGLGLGLLGLVPLPLPSKRKRDRWN"		
	<6237..8494		
	/note="similar to GenBank Accession Number M12788 (ECOCCA)"		
gene	6886..7924		
	/gene="cca"		
CDS	6886..7924		
	/gene="cca"		
	/EC_number="2.7.7.25"		
	/function="tRNA nucleotidyltransferase"		
	/note="CG Site No. 933"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89136.1"		
	/db_xref="GI:882578"		
	/translation="MKIYLVGAVRDALLGLPVKDRDWVVVGSTPQEMLDAGYQOVR DFVLPQTHQEYALARTKSGSGYGTCTAAPDVLTLEDLKRDLTLINALAQDD NEFLIDYQNGLDQNLRLRHVSFAFGEDELRLVLRVAARFAAHLGFRADETLALM REMTHAGELEHLPYRWKTESALTTRNPQVFQVLRCDCALRVLPPEIDALFGVPA PAKWHPIEDTITLMTLSMAALSPOQDVRFATLCHDLGKGLPPPELWPRHHGHPA GKLVLEOLCORLPNEIRDLARLVAEFHDLITFFPMLNPKTIKVLKFDSDANWRKPOR VEOLALTSEADVRGRTGFESADYPOGRWLREAMEVAQSVPTKAVVEAGFKGVEIREL TRRRIAASWKKQRCQKPE"		
	complement(8105..8926)		
gene	/gene="baca"		
CDS	complement(8105..8926)		
	/gene="baca"		
	/note="CG Site No. 29739"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89137.1"		
	/db_xref="GI:882579"		
	/translation="MSDMHSLTIAAILGVVVEGLTEFLPVSGTGHMIIIVGHLGFGDT AKTFEVIOQLSILAVVVMFRRRLFGILIGHFGRPLQHGEGSKGRLTLIHLLGMIPA VWLGLELHDTRIKSLFNFINVMYALVGGILLIARAECLKPKPRAPGLDDMTYRQAFMI GTCQCLALWFGFSKGATISGGLMVGSRYAASEFSFLLAVPMMGATALDLKYSWGFI LTSGDIPMFVAGFITAFVVALIAIKTLQLIKRIKSFIPFAIYRPIVAAYVYVVF"		
misc_feature			
	complement(8455..10099)		
	/note="similar to GenBank Accession Number L12966 (ECOBACTRA)"		
	complement(9016..9387)		
	/note="alternate name ygiG; ORF_fl23"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89138.1"		
	/db_xref="GI:882580"		
	/translation="MDMVFTEQLSVTTTIGVVDMEQTIQKLVDFDIEMADNFKAK SDVADCLSYADIAETVSVHVEGARFALVERVAEEVAELLLARNSFWRIKLSKPGA VARAANVGVIIRGNLKENN"		
	9489..10106		
	/note="alternate name ygiH; ORF_0205"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89139.1"		
	/db_xref="GI:882581"		
	/translation="MSAIPGCMILIAVLCGSISSAILVCRLCGLDPDPTSGSGNPGAT NVLRIIGKGAAVAVLIPDLVKMLPVMGAYELGVSPEWLGIAIAACLGHIWVPVFGF KGGKVATAFGATAPIGWDLTGVMAGTWLLTVLLSGYSSIGAIVSALIAPIYVWKEFP QTFPPVMSLCLILLRHNDIQRWRQETKIWTKFKRREKDE"		
	complement(10119..11051)		
	/note="alternate name ygiP; ORF_310"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA89140.1"		
	/db_xref="GI:1203799"		
	/translation="MLNSWPLAKDLQVLVEIVHSGFSAAATLGQTPAFVTKRIQIL ENTLATTLLNRSARGVALTESGQRCYEHALEILTQYRLVDDVDTQIKRPEGMIRIC SFGKSHRIAPATELMRNYPLOQSFELDRQIDLVDQNDIDLRINDEIDPYIAH LLTKNRILCAAPYLOKYPQOSLOELSRHDCLVTKERDWTGHWELNGOEKKSVK VSHLSSNSGIVLOWALEGKGLMRSEWDVLPLESGKLVQVLPYVAQSANIWAVYR EPYRSMKLRVCEVFLAANCQRLCKPDEGYQM"		
	10928..11584		
	/note="similar to GenBank Accession Number M16194 (ECORPSU)"		
	10978..11584		
	/note="similar to GenBank Accession Number L14781 (ECOTATTDB)"		
gene	11258..11269		
	/gene="ttdA"		
CDS	11258..11269		
	/gene="ttdA"		
	/EC_number="4.2.1.32"		
	/codon_start=1		
	/transl_table=11		
	/product="tartrate dehydratase"		
	/protein_id="AAA89141.1"		
	/db_xref="GI:882584"		
	/translation="MMSESNKQAVNKLTEIVANFTAMISTRMPPDDVVDKLOKDAE EATVKAPLRINAVEIFDEVNTKNGTSGVFWVWTDIIPONDAAIEVYIMAGGCTLPG RSKVLPSGEYGVYKFNISITLAVNACPPVPLVGVGIATSVETAVALSKAILRPI GSRHPNPKAALELRLEGLNRLGIPQGLTGSSVMGVHIESAARHPSTIGVAVSTG CWARRRCTLHADLTLENLSHRSAL"		
	12166..12771		
gene	/gene="ttdB"		
CDS	12166..12771		
	/gene="ttdB"		
	/EC_number="4.2.1.32"		
	/codon_start=1		
	/transl_table=11		
	/product="tartrate dehydratase"		
	/protein_id="AAA89142.1"		
	/db_xref="GI:882585"		
	/translation="MKKILTPTKAEDLQDIRVGDVILYLTGLVTCRDVCHRRRIELK RPYPDLNGKAIFHAGPIVRKNGDKWEMSVGPTTSMRMESFEREFTEQTCGLVVGK GGMPLTEGCKQFKALHVIFFPAGCAVLAATQVDEIEEVHVTGLMPELWCVKRVKEF GGLIVSDTHGNLLIAENKKLFAERRDPVIEEICEHVHIK"		
	12819..14282		
	/note="alternate name ygiE; ORF_0487"		
	/codon_start=1		


```

/db_xref="GI:12517626"
/translation="MILFADYNTPTLPFAISFVLVLLIGLLEIFALICGMLSGALDAHLID
HYDSITTHCHISOALHNLIGRLPALVLCILLAGFEGILIGLLOHTCIMVQSPISLNF
VVPVSLHTIIVAVHYTKIVAPWIPROHSSAITEEYIGSMALITGHQATSGNCECK
LTDHFGQIHVLLLEPEBGKIFTKGDKVLIICRLSATRYLAENNP"
/gene="24403"
/length=3838
/feature="putative membrane: Not classified"
/note="Residues 1 to 553 of 553 are 98.73 pct identical to
residues 1 to 553 of 553 from Escherichia coli K-12 Strain
MG1655: B3051"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAG58185.1"
/db_xref="GI:12517627"
/translation="MDDIVNSVPSMNTAIIAVCILFIIGIIFARLYRRASAEQAFVR
TGLGGQVVMGGGAIVPIFHEIIPINMNTLKLEVSRTIDSLTKDRMRVDVVVAF
VRVPSVEGTATAQTLCQRTLSPEDRLMLVEDKFVDALRATAQMTMHELODTRRENF
VOGNTVAEDLSKNGLESVSLTNEFNOTSKHFNNAFDAEGLTKLQDETERRRR
ERNEEDVAVREKRDALSKLELEQOEAFMTLEQEQVKTREAOAKIAFEA
ERRREASQTRILAEQIQTETIDREAVRSKVEAREVRIKELEQQOVTETIANQTS
IAIAKSEQSQSARANLALAEAVSAQONVETTPQTAADRAKQVALIAAQADET
AVELTVRAKAEKEAEMQAAIVELAEATRKKGIAEAEQALNDAINVLSDQTSLK
FKLALQALPAVIEKSVPEPMKSIDIKIIQVDGLNRGTAGDANTGNVGGNLAEQAL
SAALSYRQALPILDSLLNEIGVSGSLAALTSLTPTTPVAENVE"
/length=5799
/note="O-island #125: Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(6293..7726)
/gene="24405"
/complement(6293..7726)
/gene="24405"
/function="putative enzyme: Not classified"
/note="Residues 1 to 477 of 477 are 99.58 pct identical to
residues 1 to 477 of 477 from Escherichia coli K-12 Strain
MG1655: B3052"
/codon_start=1
/transl_table=11
/product="putative kinase"
/protein_id="AAG58186.1"
/db_xref="GI:12517628"
/translation="MKVTLPEFERAGVVMGVDMLDVWYGPSSRSISPEAPVPVKVN
TIERPGGAANVANMIASLGANLVGLTGIDDAARALSKSLADVNVKDPVSVPTHP
TITKRLVSRNOQLIRLDFEGFEGVDFQPLHERINQALSSIGALVLSYAKALASV
QOMTQLARKAGVPLIDPKGDFERYRGATLLTPNLSEFAVVGKCTEEIEIVGMK
LIADYLSALLVTRSEQMSLLQPKAPLHMTQAOEYVDVGADTVIGVIAATLAA
GNSLEEAFFANAAAGVVGKLGSTVSPLENAVRADTGFVGMTEELKLVAVP
ARKRGEVMTNGVDFILHAGHSYLANARKLDRLIVAVNSDASTKRLKGDSPVNP
LEORMIVLGALEAVDWPVSEEDTPQRLIAGILPDLVLKGGDYKPEETAGSKYEWANG
GEVLVLFEGSCSTNIKKIQQKKG"
/complement(7774..10614)
/gene="24406"
/complement(7774..10614)
/gene="24406"
/function="enzyme: Macromolecule synthesis, modification:
Proteins - translation and modification"
/note="Residues 1 to 946 of 946 are 98.73 pct identical to
residues 1 to 946 of 946 from Escherichia coli K-12 Strain
MG1655: B3053"
/codon_start=1
/transl_table=11

```

Alignment Scores:

Pred. No.:	5,18e-137	Length:	12088
Score:	2385.00	Matches:	475
Percent Similarity:	99.79%	Conservative:	1
Best Local Similarity:	99.58%	Mismatches:	1
Query Match:	99.67%	Indels:	0
DB:	1	Gaps:	0

```

US-09-912-020-325 (1-477) x AE005534 (1-12088)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValClyAspValMet 20
DB 7726 ATGAAGTAAGCTGCCAGAGTTTGAAGCTGCAGAGTATGCTGCTGATGTGATG 7667
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 7666 CTGGATCGTTACTGTGTCAGGCCCTCCAGTCTGATCTCGCGGAAGCGCGTGGCG 7607
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
DB 7606 GTTAAAGTGAATACCATCGAAGACGTCGCGCGCGCGGCTAACGTGCGGATGATC 7547
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArg 80
DB 7546 GCTTCTCTCGGTGCTAATGCACGCTGTCGGGTTCAGCGGCATTGACATGACGCGC 7487
QY 81 AlaLeuSerLysSerLeuAlaAspValAlaAsnValLysCysAspPheValSerValProThr 100
DB 7486 GCGCTGAGTAATCTCTGCGCGACGTCACAGTCAATGCGACTTCGTTCTCTACCGAG 7427
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
DB 7426 CATCGGACCATTACCAAAATTACGGGTACTTTCGCCCAACCAACAGCTATCGCTGAT 7367
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
DB 7366 TTTGAAGAAGGTTTCGAAGGTTCATCGCGCGCTGCACGCGGATTAATCAGGCG 7307
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
DB 7306 CTGAGTTTCGATTGGCGCGCTGCTGCTTCTGACTAGCCAAAGGTCGCTGCAACGCTA 7247
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
DB 7246 CAGCAGATGATCCAACTGCGCGCTAAAGCGGTGTTCCGGTCTGCTATTCATCAAAAGT 7187
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
DB 7186 ACCGATTTTGAGCGCTACCGCGCTTCAGCTGTGTTAAACGCCGAATCTCTCGAAATTTGA 7127
QY 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
DB 7126 CTGTTGTCGTTAATGTAAGCCCAAGAGATTTGACCGCGGATGAACTGAT 7067
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlnMetSerLeuLeu 240
DB 7066 GCCGATTACGAACCTCTCGCTCTGTTAGTACCCCGTTCGGAACAGGATGTCGCTGCTG 7007
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
DB 7006 CAACCGGTTAAAGCGCGCTGCATATGCCAATCAGCGCGAGNAAGTGTATACGCTTACC 6947
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
DB 6946 GGTGCGCGCGACACCGTGAATGCGCTCTGCGCGCAACGCTGCGACGCGGTAAATTCGCTG 6887
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValClyLysLeuGlyThr 300
DB 6886 GAAGAAGCCTGCTCTTTGTCCAATCGCGCGCTGTTGTTGTTGTCGCGCAAACTGGGAGC 6827
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
DB 6826 TCCACGGTTTCGCCGATCGAGCTGGAACCGCTACGCTGGACGCGCGCATACCGCTTT 6767
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
DB 6766 GCGGTGATGACCGAAGAGAACTGAAGCTGCTGAGTGGCAGCGCGTAACGCTGGTGA 6707
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360

```



```

/gene="ECs3913"
/note="probable ferrichrome-binding protein, similar to
ferrichrome-binding proteins e.g. [Bacillus subtilis]
gi|585132|sp|P37580|FHUB_BACSU percent identity 27 in 220
aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative ferrichrome-binding protein"
/protein_id="BAB37336.1"
/db_xref="GI:13363386"
/translation="MRLPFSLLILSFARATEPVQVFTDGLGRVTVPAHPKRVLSL
HDLTITLIEIGVPPVASHGRTPDGSFIRSGALLTGVPDINSIAFETADIDIE
AIVAAKPLDITETPTPIERLEIAPTVIDHLKGAPETIKRLAELTGTQSOLAI
LERRQAQINALKATLDQKIVSVIQANQKINVMHVSYHSLGRVLRDAGFRFPPLIE
SIPEGGRMDVSAERLELDADFVATWRGDTGGKPDDELAAMEKVMYWCQFLTACRS
GRYVLISREALISNSFASGLMAAQISQIAGRPLPEAK"
complement(2885..3934)
/gene="ECs3914"
complement(2885..3934)
/note="ECs3914"
/note="probable ferrichrome ABC transporter
(permease), similar to ferrichrome ABC transporter
(permease) e.g. [Bacillus subtilis]
gi|1706797|sp|P49937|FHUG_BACSU percent identity 33 in 319
aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative permease of ferrichrome ABC
transporter"
/protein_id="BAB37337.1"
/db_xref="GI:13363387"
/translation="NSRIMNVCLRLRVKSTLYRPNKLVLLGLFLFAVGILIFGL
MHGSPFVASEVGRALFAPENVSTDAIRYIQDRLPRVIMALLCGAMLMGAAQMSI
ARNGLADPLGIVGKCCSVAVLWLIQFPMGLGWRPVRVAGLGGVLVALIVFCAREI
SRPRVLIGIVSFFAGVFMFTADVDQATLMMLSGLSHAANMLVIGISACWM
LPALLLHTARTADIALIGHQVATGLGVNSRLALLRVAAPILITAVCVSCVNGIF
VGLIAPHISRFLLRGQTTLLGSVAGSALLVILADSIGRLAFLQLPAGIISLIG
GPFFLLLRNRP"
complement(3931..4914)
/gene="ECs3915"
complement(3931..4914)
/note="ECs3915"
/note="probable ferrichrome ABC transporter
(permease) similar to ferrichrome ABC transporters
(permease) e.g. [Synechocystis sp.] gi|7442493|pir||S74438
percent identity 43 in 315 aa, [Bacillus subtilis]
gi|1706795|sp|P49936|FHUB_BACSU percent identity 39"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative permease of ferrichrome ABC
transporter"
/protein_id="BAB37338.1"
/db_xref="GI:13363388"
/translation="NMRTGLTVIFLALLVGVVYPGIGARFIAPQTVLAQFLHPEP
QNFQHNIVRLRLPLRAALLTGASLVAGALLOAVIRNPLGEPIHILGNAGAALAVY
RAASGLAFPVGRPLLASTGGALLFLILLSSAGRSLTPMKVTLGCVALSASFYSSI
TAAIILLDEQILNMRWTLADQADWATIGTSWFSIGGVVLAIIYLAPSLNMILALG
DMAAGLVSVLRTFTTLLAIALICGAASVAGTGFVGLLVPOIVRLVSADLRVL
LPLSACVGAALLLADIARTLFTPHLATGYMTALVGAPVFVIMATRMFK"
complement(4911..5720)
/gene="ECs3916"
complement(4911..5720)
/note="ECs3916"
/note="probable ABC-type iron-siderophore transport system
ATP-binding protein, similar to ABC-type iron-siderophore
transport system ATP-binding proteins e.g. [Synechocystis
sp.] gi|7442509|pir||S74440 percent identity 52 in 248 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11

```

```

/product="putative ABC-type iron-siderophore transport
system ATP-binding protein"
/protein_id="BAB37339.1"
/db_xref="GI:13363389"
/translation="MAHNKAGGGLILDNVVSAGYHKRIIVDGVSFSPVPTKMTVLVCAN
CGKSTLLSTIARVLPMGGSLLDKALHEQPTKALSRRLGLPQSLPLPGLTVYIE
LVSRGRFPQNFIRQMSDADEQAEALKTCTQEFAPHLPEKLSGGQRQRCWIAVL
AQKTPYILLDEPTWLDLRQVEILELHDLTKRHGRIVVYVLHDLNFAVNTGDTLIF
LRQGGKVRVRLNEGECHTPELVKAIVEDVDVHASINTLTGKPFPMFPRGVEKV"
Alignment Scores:
Pred. No.: 2,59e-135 Length: 307962
Score: 2385.00 Matches: 475
Percent Similarity: 99.79% Conservatives: 1
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 99.67% Indels: 0
DB: 1 Gaps: 0
US-09-912-020-325 (1-477) x AP002564 (1-307962)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
|||||
Db 24982 ATGAAGTAACTGCTGCCAGAGTTTGACGTGCAGGAGTGTGATGTTGCGTGTGATG 24923
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
|||||
Db 24922 CTGGATCGTTACTGTGTCACGGCCCTCCAGTCGTATCTCCCGCAACCGCGCGTGC 24863
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
|||||
Db 24862 GTTAAAGTCAATACCATCGAAGAACGTCGGCGCGCGGCTAACGTGGCGATGATATC 24803
QY 61 AlaSerLeuGlyAlaAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
|||||
Db 24802 GCTTCTCTCGGTCTAATGCACGCCCTGGTGGGTTCAGCGGCATTCACGATGACGCGCC 24743
QY 81 AlaLeuSerLysSerLeuAlaAspValAlaAsnValLysCysAspPheValSerValProThr 100
|||||
Db 24742 CGCCTCAGTAAATCTCTGCGCGACGTCAAGTCAATGCGACTTCCTTTCTGTACCGACG 24683
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
|||||
Db 24682 CATCCGACCATTAACCAATATACGGGTACTTTCCCGCAACCAACAGCTGATCCGCTCTGGAT 24623
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
|||||
Db 24622 TTTGAAGAGAGTTTCCAGAGTGTTGATCCGACCGCGCTGCACGAGCGGATTAATCAGCGC 24563
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
|||||
Db 24562 CTGAGTTTCGATTGGCGCGTGGTGTCTTCTGACTACGCCCAAGGTCCGCTGGCAAGCGTA 24503
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuLeuAspProLysGly 180
|||||
Db 24502 CAGCAGATGATCCAACTGCGCGCTAAAGCGGGTGTTCGCGGTGCTGATTGATCCAAAAGGT 24443
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
|||||
Db 24442 ACCGATTTTGAGCGCTACCGCGCGCTACGCTCTTAACGCCGGAATCTCTCGGAATTTGAA 24383
QY 201 AlaValAlaGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
|||||
Db 24382 GCTGTGTTCGGTAAATGTAAGACCGAAGAGAGATTGTTGAGCGCGCATGAAACTGATT 24323
QY 221 AlaAspTyrGlnLeuSerAlaLeuLeuValThrArgSerGlnGlnGlyMetSerLeuLeu 240
|||||
Db 24322 GCGGATTAACAACTCTCGGCTCTGTTAGTACCCCGTTCCGAAACAGCGTATGTCGCGTCTG 24263
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
|||||
Db 24262 CAACCGGGTAAACCGCGCGTGCATATGCCAACTACGCGCAGCAAGAACTGTATGACGTACC 24203
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280

```



```

|||||
Db 24202 GGTGGCGGACACGGTGATTGGCGTCTCTGGCGCAACGCTGGCGAGCGGGTAATTCGCTG 24143
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValGlyLysLeuGlyThr 300
Db 24142 GAAGAAGCCTCTCTTTGGCAATGCGGCGGCTGGTGGTGGTCGCAACATCGGGAGC 24083
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 24082 TCCACGGTTTCGCCGATCGAGCTGGAAACCTGCTAGTGACCGCGCAGATACCGGCTTT 24023
Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 24022 GGCCTGATGACCGAAGAGGAACTCAAGCTGCCTAGTGGCAGCGCGTCAACGTTGGTAA 23963
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 23962 AAAGTGGTAATGACCAACGGTGTCTTCGATATCTACATGCGGCGCAGCTCTCTATCTG 23903
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 23902 GCAATGCCCGCAGCTGGGTGACCGCTGTGTTGCTGCAACAGCGATGCCCTCCACC 23843
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 23842 AAACGGCTGAAGAGGGATTCCCGCGCGGTAAACCCACTCGAACACGCGTATGATTGCTG 23783
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 23782 GCGGACATGGAAGCGGTGCTGCTGGTGTGCTGTTGAAGAAGACACGCCCGCAGCTG 23723
Qy 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyLysAspTyrLysProGluGluIle 440
Db 23722 ATCGCGGGATCCTGCCAGACTCTGCTGGTGAAGCGGTGACTACAAACCGAGAGATC 23663
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 23662 GCCGGAGTAAAGAAGTCTGGCCCAACGGCGCGCAAGTGTGTGGTGTCAACTTTGAAGAC 23603
Qy 461 GlyCysSerThrAsnIleIleLysLysIleGlnGlnAspLysLysGly 477
Db 23602 GGTTCGTCGACCACTAATATCAAGAAGATCCACAGGATAAAAAAGGC 23552

RESULT 7
AF155126 Salmonella typhimurium 1431 bp DNA linear BCT 20-MAR-2001
LOCUS AF155126
DEFINITION Salmonella typhimurium ADP-heptose synthase (rfab) gene, partial
cds.
ACCESSION AF155126
VERSION AF155126.1 GI:13383335
KEYWORDS
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE 1 (bases 1 to 1431)
AUTHORS Jin, U.-H., Chung, T.-W. and Kim, C.-H.
TITLE ADP-heptose synthase (rfab) gene of salmonella typhimurium
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1431)
AUTHORS Jin, U.-H., Chung, T.-W. and Kim, C.-H.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1999) Oriental Medicine, Dongguk University,
Sukjang-Dong 707, Kyungju 780-714, Korea
FEATURES
source
1. .1431
/db_xref="taxon:602"
/organism="Salmonella typhimurium"
1. .>1431
/gene="rfab"
/feature="waab"
1. .>1431
/gene="rfab"

```

```

/codon_start=1
/transl_table=11
/product="ADP-heptose synthase"
/protein_id="AAK20933.1"
/db_xref="GI:13383336"
/translation="MKVTLPFERAGVMVGVDMVDLYWYGPCTCRISPEAPVPVKVN
TVBERPGGAANVAMNIAISGANARLVLTGIDDAARALSKTLAEVNVKDFVSVPTHP
TITKRLVSRNOOLIRLDFEGFEGVDPOPLHERINOALGSGIALVLSYAKGALTSTV
QTMISLARQAGYPLVDPKGTDFERYRGATLLTPNSEFENAVAGKCKSEDELVERGMK
LTADYLSALLVTRSEGGMTLLQPNKAPLHMTQAOEYDVTGAGDTVIGVLAATLAA
LNDEEACAFANAAAGVVGKLTSTVSPIELENAVGRADTGFVGMVTEEDRQAAS
ARRKGEKVVTNGVDFIDHAGVSYLANARKLDRLIVAVNSDASTKRLKGESRPVNP
LEORMTVLGALESVDWVSFEEDTPORLIAGILPDLLVRKGGDYKPEIAGSEEVWANG
GEWMLNFDGCGSTTNLIKIOIESEK"
BASE COUNT 311 a 357 c 467 g 296 t
ORIGIN
Alignment Scores:
Pred. No.: 1.95e-131 Length: 1431
Score: 2276.00 Matches: 447
Percent Similarity: 97.90% Conservative: 19
Best Local Similarity: 93.91% Mismatches: 10
Query Match: 95.11% Indels: 0
DB: 1 Gaps: 0
US-09-912-020-325 (1-477) x AF155126 (1-1431)
Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValGlyAspValMet 20
Db 1 ATGAAAGTAAACACACGTCAGAGTTCGAACGCGCAGAGGATATGTTGGTGGTGGTATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGTATCGCTATGTTATGTCGCCACATGCGGTATTTTCCCGGAAGCGCGGTGCCGTG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAGGTAATACCGTTGAGGAACGCCGCGCGCGCGCAACGTCGGCATGAACATT 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCGTCTCTGGAGCGAAGCGCGCTCTGGTGGCTGACGGGTATTTGATGACGCGCGCGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGAGCAAAACCGCTGGCGGAGTCAATGTGAAGTCGCGACTTCCTTCTGTCGCCAGC 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACGATTACCAAACTGCGCGTACTATCACGTAATCAGCAGCTCATTCGCTTTGAT 360
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTTGAAGAAGCTTTTGAAGCGCTGGACCGCGCGCGCTTGCATGACGCTATCAACACGCGC 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAspVal 160
Db 421 CTGGATCGATCGCGCGCTGGTATTTCCGATTATGCCAAAGCGCTCTGACTAGCGGTG 480
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGACTATGATTTCCTTGGCGCGCGCGCGCGCGCGCTGCTCATCGATCCGAAAGGA 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACGGATTTTGAACGTTTACCGCGCGCGCGCGCTGCTGACGCGCAACCTTCTGAAATTTGAG 600
Qy 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCGTCCGCGGGAANTGTAAAGCGAAGACGAACTGGTTGACGCGCGCATGAACCTATT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
|||||

```



```
protein_bind
/translation="MASTYRTTIRANTYOFRETTMIDPKKIBQIARQVHESMPKGIRE
FQEDIEKKIQTQLQSLRLDLVSREEFDVQTVLLTRKILALLEQRLSELEARDKP
BEVGPAPAPPPVDPOQE"
complement(1579..1797)
/gene="glgS"
/note="glgS"
/note="STM3197"
complement(1579..1788)
/gene="glgS"
/note="glgS"
/note="similar to E. coli glycogen biosynthesis, rpos
dependent (AAC76085.1); Blastp hit to AAC76085.1 (66 aa),
78% identity in aa 2 - 66"
/codon_start=1
/transl_table=11
/product="rpos dependent glycogen biosynthesis protein"
/protein_id="AAL22071.1"
/db_xref="GI:16421754"
/translation="MLMNNVYSLNFFDLARSFARMQAGRPVDIQAVTGNMDEEH
RWFCKRYALYCCQATQAKRLEH"
complement(1792..1797)
/gene="glgS"
/note="putative RBS for glgS; RegulonDB: STMSIH003228"
complement(1885..1905)
/note="putative binding site for CRP, RegulonDB:
STMSIH000051"
/bound_moiety="CRP"
2033..2666
/gene="STM3198"
2033..2038
/gene="STM3198"
/note="STM3198"
2046..2666
/note="putative RBS for STM3198; RegulonDB: STMSIH003229"
/note="STM3198"
/note="similar to E. coli putative oxidoreductase
(AAC76086.1); Blastp hit to AAC76086.1 (209 aa), 72%
identity in aa 1 - 201"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22072.1"
/db_xref="GI:16421755"
/translation="MTLFAEYNSPYLFAIAFFVFFGVLEMISLIFGHFLSGALDAHL
HYDALSSGAGALHYLNIGRYPALVLCGLAGYGLGILQHGGIMLWQPLSNLL
LYPLSLVSVFVHYSGKILAPLWPRDESSALREEFITGGMAITGHRAVAGTPCEGK
FTDKFGQIHYLLPEKKGKEFKGDKVLIVCRLSATRYLAERTFYV"
2671..4364
/gene="yqik"
/note="STM3199"
2671..2676
/gene="yqik"
/note="putative RBS for yqik; RegulonDB: STMSIH003230"
2685..4364
/gene="yqik"
/note="yqik"
/note="similar to E. coli putative membrane protein
(AAC76087.1); Blastp hit to AAC76087.1 (553 aa), 92%
identity in aa 1 - 553"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAL22073.1"
/db_xref="GI:16421756"
/translation="MDDVFGILPSWMFTAIVAVIVLLIIGIIFARLYRRASQAFVR
TGLGGQVSGVSGAIVMPFHEIIPINMTLKVESRATVDSLITKDRMRVDVWVAF
VRVKSVEGIAAOTLGORTLSPEDLRMLVEDKFVALRATAOQMTHELODTRENF
VOGVNTVAEDLSKNGLESLVFNQTSKEHENPNNAFDEAGLTKTOETERRR
ERNEVDQVEAVREKRDALERKLEIEQOEAFMLEQQQVKTRTAQGNAKIAFEA
EHRBAEQTRILAEQIOETIEREQAVRKRVEERVIKEIEQQQVQVTAQNTKS
IAIAAKSQEAEALQAAIIELAERATKRAQONVETTRQTAEDRAKQVALTAAQDAETK
AVELTVRAKAEAEALQAAIIELAERATKRAQONVETTRQTAEDRAKQVALTAAQDAETK
FALLOLOSAPVIEKSVPMKSIDGKIIOVDGLNRGATAGDVAAGGANGNLAEQAL
KALIVTQAPLIDSLINEIGIAGGSLKALTTPLYSSATDEINREAVIEQ"
complement(4824..6269)
/gene="rfae"
/note="STM3200"
gene
/translation="MASTYRTTIRANTYOFRETTMIDPKKIBQIARQVHESMPKGIRE
FQEDIEKKIQTQLRDLVSREEFDVQTVLLTRKILALLEQRLSELEARDKP
BEVGPAPAPPPVDPOQE"
complement(1579..1797)
/gene="glgS"
/note="glgS"
/note="STM3197"
complement(1579..1788)
/gene="glgS"
/note="glgS"
/note="similar to E. coli glycogen biosynthesis, rpos
dependent (AAC76085.1); Blastp hit to AAC76085.1 (66 aa),
78% identity in aa 2 - 66"
/codon_start=1
/transl_table=11
/product="rpos dependent glycogen biosynthesis protein"
/protein_id="AAL22071.1"
/db_xref="GI:16421754"
/translation="MLMNNVYSLNFFDLARSFARMQAGRPVDIQAVTGNMDEEH
RWFCKRYALYCCQATQAKRLEH"
complement(1792..1797)
/gene="glgS"
/note="putative RBS for glgS; RegulonDB: STMSIH003228"
complement(1885..1905)
/note="putative binding site for CRP, RegulonDB:
STMSIH000051"
/bound_moiety="CRP"
2033..2666
/gene="STM3198"
2033..2038
/gene="STM3198"
/note="STM3198"
2046..2666
/note="putative RBS for STM3198; RegulonDB: STMSIH003229"
/note="STM3198"
/note="similar to E. coli putative oxidoreductase
(AAC76086.1); Blastp hit to AAC76086.1 (209 aa), 72%
identity in aa 1 - 201"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22072.1"
/db_xref="GI:16421755"
/translation="MTLFAEYNSPYLFAIAFFVFFGVLEMISLIFGHFLSGALDAHL
HYDALSSGAGALHYLNIGRYPALVLCGLAGYGLGILQHGGIMLWQPLSNLL
LYPLSLVSVFVHYSGKILAPLWPRDESSALREEFITGGMAITGHRAVAGTPCEGK
FTDKFGQIHYLLPEKKGKEFKGDKVLIVCRLSATRYLAERTFYV"
2671..4364
/gene="yqik"
/note="STM3199"
2671..2676
/gene="yqik"
/note="putative RBS for yqik; RegulonDB: STMSIH003230"
2685..4364
/gene="yqik"
/note="yqik"
/note="similar to E. coli putative membrane protein
(AAC76087.1); Blastp hit to AAC76087.1 (553 aa), 92%
identity in aa 1 - 553"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAL22073.1"
/db_xref="GI:16421756"
/translation="MDDVFGILPSWMFTAIVAVIVLLIIGIIFARLYRRASQAFVR
TGLGGQVSGVSGAIVMPFHEIIPINMTLKVESRATVDSLITKDRMRVDVWVAF
VRVKSVEGIAAOTLGORTLSPEDLRMLVEDKFVALRATAOQMTHELODTRENF
VOGVNTVAEDLSKNGLESLVFNQTSKEHENPNNAFDEAGLTKTOETERRR
ERNEVDQVEAVREKRDALERKLEIEQOEAFMLEQQQVKTRTAQGNAKIAFEA
EHRBAEQTRILAEQIOETIEREQAVRKRVEERVIKEIEQQQVQVTAQNTKS
IAIAAKSQEAEALQAAIIELAERATKRAQONVETTRQTAEDRAKQVALTAAQDAETK
AVELTVRAKAEAEALQAAIIELAERATKRAQONVETTRQTAEDRAKQVALTAAQDAETK
FALLOLOSAPVIEKSVPMKSIDGKIIOVDGLNRGATAGDVAAGGANGNLAEQAL
KALIVTQAPLIDSLINEIGIAGGSLKALTTPLYSSATDEINREAVIEQ"
complement(4824..6269)
/gene="rfae"
/note="STM3200"
CDS
complement(4824..6257)
/gene="rfae"
/EC_number="2.7.-.-"
/note="bifunctional; similar to E. coli putative kinase
(AAC76088.1); Blastp hit to AAC76088.1 (477 aa), 93%
identity in aa 1 - 476"
/codon_start=1
/transl_table=11
/product="putative sugar nucleotide transferase domain of
ADP-L-glycero-D-manno-heptose synthase"
/protein_id="AAL22074.1"
/db_xref="GI:16421757"
/translation="MKVNLPAFERAGVMVGVDMILFRWYWGPTCRISPEAPVVKVN
TVEERPGGAANVAMTIASLGANARVLGLTGIDDAARALSKTLAEVNVKQDFVSPVTHP
TTTKLRVLSRNOQLRLDFEGFEGVDPQPLHERHTNOALGSGICALVLSYAKGALTSV
QPMISLARQAGVPLIDPKGTDFERYRGATLTPNLSEFEAVAGCKSSEDELVERGMK
LIADLDSALLVTRSEQMTLLQPNKAPLHMFTQAGEYVDYVTGAGDGYTGLVLAATLAA
GNTDEACVFNAAAGVGVKIGTSTVPIELENAVRADTGFVGMTEEBEERQAVAS
ARKRGEKVMYTNQVDFILHAGVSYLANARKLIDRLIVAVNSDASATKRLKGSRPVNP
LEQRMITVLGALESVDVWVSFEEDTPQRLIAGILPDLILVKGGDYKPEIAGSEVWANG
GEVMVLFEDGCGTTNIIKKIOTSEK"
complement(6264..6269)
/gene="rfae"
/note="putative RBS for rfaE; RegulonDB: STMSIH003231"
complement(6305..9163)
/gene="glnE"
/note="STM3201"
complement(6305..9148)
/gene="glnE"
/EC_number="2.7.7.42"
/note="regulates p-II (GlnB) and GlnK; similar to E. coli
adenylylating enzyme for glutamine synthetase
(AAC76089.1); Blastp hit to AAC76089.1 (946 aa), 87%
```

Alignment Scores:

Pred. No.:	2,33e-129	Length:	20604
Score:	2265.00	Matches:	445
Percent Similarity:	97.48%	Conservative:	19
Best Local Similarity:	93.49%	Mismatches:	12
Query Match:	94.65%	Indels:	0
DB:	1	Gaps:	0
US-09-912-020-325 (1-477) x AE008847 (1-20604)			
Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValGlyAspValMet 20			
Db 6257 ATGAAAGTAATCTGCCAGCGTTTGAACGTGCGAGCGTCAATGTTGGGTGATGTAATG 6198			
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40			
Db 6197 CTGGATCGCTATTGTTATGCGCCCACTTGCCTGTTTACCGGAAGCGCGGTGCCGCTG 6138			
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60			
Db 6137 GTTAAGGTAATACCGTTTGAAGAACGCCGCCGCGCGCGCGGCGGCGATGAACATT 6078			
Qy 61 AlaSerLeuGlyAlaAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80			
Db 6077 GCGTCTCTGGGACGCAACGCCCGCTCTGGTGGCTGACGGGTATTGATGACGCCGCCGCGC 6018			
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100			
Db 6017 GCGCTGAGCAAAACGTCGGCGAGGTCAATGTGAAGTGCACATTCGTTTCTGTGCCGACG 5958			
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120			
Db 5957 CATCCGACGATTACCAAACTGCGCGTACTATCACGTAATCAGCAGCTCATTCGCTTGAT 5998			
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140			
Db 5897 TTTCAAGAGCGCTTTCAGGGCGTGGACCCGCCCGCTGATGAGCGTATCACACGCGCG 5838			
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160			

Db 5837 CTGGGATCGATCGCGCGCTGTTATGTCGATATGCAAAAGCGCTGTACACGCGTG 5778
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuLeuAspProLysGly 180
Db 5777 CAGACATGATTCCTCGCGCGCCAGCGCGCGTGCCTCATCGATCCGAAAGGA 5718
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 5717 ACGGATTTGAACGTTACCGCGCGCCAGCTGCTGACGCCAAACCTTCTGAATTGAG 5658
Qy 201 AlaValAlaGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 5657 GCGGTGCGCGGGAATATAAAGCGAAGACGAACGAGTGTGAACGCGCATGAACCTCAT 5598
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 5597 GCTGATTAACGACCTTTCGCGCTGTTGTCACGCGCTTCCGAACAGGAAATGACGCTGCTG 5538
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 5537 CAACCGAATAAGCGCGCTACATATGCGCAGCGCAGCGCAGGATTTATGATGTAC 5478
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 5477 GGTGCGGCGGATACGATGATCGCTGCTGCGCGGACGCTGCGCGCGGAAATACCGTG 5418
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
Db 5417 GAAGAGCGCTGTATTTCCCAATCGCGCGCGCGCTAGTGGTAAATCCGGGACG 5358
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 5357 TCACCGGTTCCCTATTGAGCTGGAACAGCAGTGCAGCGCGCGGATACCGGCTTC 5298
Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 5297 GCGGTATGACCGAAGAGGATGTAGACAGCGCGCTGCGCGCGCGCGTAAAGCGTGGCGAG 5238
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 5237 AAAGTGGTTCATGACCAACGCGCTTTTCGATATTCGACGCGCGGCGCATGCTCTATCTG 5178
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 5177 GCGAAGCGCGCAACTGGCGCACCGCTGATGTGCGGTCAATAGTACCGCTCGACT 5118
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 5117 AAACGCGCTGAAGCGCAAGCGCTCGGTTAATCCGCTCGAAGCGGTATGATCGTGCTA 5058
Qy 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420
Db 5057 GCGCGCTCGATCGGTCGCTGACGTGGGTGCTCTTTGAAGAGGATACGCGCAACGACTG 4998
Qy 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyAspTyrLysProGluGluIle 440
Db 4997 ATTGCGGATTTCTGCGCGATCTGCTGTTAAAGCGCGGCGACTATAAGCGCGGAGAGATC 4938
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 4937 GCGGCGAGCAAGAGGTCTGGGCAACGCGCGGTGAAGTCAATGCTGCTGAACATTCGAAGAT 4878
Qy 461 GlyCysSerThrThrAsnIleLysLysLysIleGlnAspLysLys 476
Db 4877 GGTGTTCACGACCAATATCATCAAAAAGATCCAGACGCGAGCGAG 4830
RESULT 9
AL627278/c 258050 bp DNA linear BCT 06-JUN-2002
LOCUS AL627278
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome; segment 14/20.
ACCESSION AL627278 AL513382

AL627278.1 GI:16504263
Salmonella enterica subsp. enterica serovar Typhi.
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
1 (bases 1 to 258050)
Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrell, B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)
21534947
11677608
2 (bases 1 to 258050)
Parkhill, J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/s-typhi/).
Location/Qualifiers
1. 258050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
69. 1544
/gene="STY3364"
/note="tolc"
69. 1544
/gene="STY3364"
/note="Similar to Salmonella enteritidis outer membrane
protein TolC precursor tolC SW:TOLC_SALEN (Q54001) (491
aa) fasta scores: E(): 0.99.0% id in 491 aa, and to
Escherichia coli outer membrane protein tolC precursor
tolc or mtcb or muka or refi SW:TOLC_ECOLI (P02930) (495
aa) fasta scores: E(): 0.89.7% id in 495 aa
Orthologue of E. coli tolC (TOLC_ECOLI); Fasta hit to
TOLC_ECOLI (495 aa), 90% identity in 495 aa overlap"
/codon_start=1
/transl_table=11
/product="outer membrane protein TolC precursor"
/protein_id="CAD07712.1"
/db_xref="GI:16504263"
/db_xref="SPTREMBL:Q823N8"
/translation="MQMKLLPLILIGLSLSPSTLSOAEMLMQVYQOARLSNPRLRKS
AADRAAFEKINEARSPILLPOLGLGADTYTNGYRDANCINSNETSASLOLTQTLFDM
SKWRGLTQERAAQIDVTYQDQTLILUNAIYFVLIADIVLSYTSYQAEIYRQ
LDQTTQRFNVLGVAITVDQNAEQYTDVLANEVTRARNLDNAVEELRQVYNYPELA
SLNVHFHTDKPFAVNAIKKEARENLSLLQARLSQDLAREIQRDQGLPTLNLTA
STGISDTSYSGSKTNSAOYDDSNMGQKIGLNFSLPLYOGGMVNSQVQAOYNFVGAS
EQLESARHSVYQTVRSFNNINASSINAYKOAVSQAQSSLDMAEAGYSVCTRTIVD
VLDAITTLIDAKQQLANARYTYLINQIKYALGTLNEQDLALNSTLGKPIPTSPES
VAPETPEQDAADAGYNAHSAAPAVQPTAARANSNNNGNPRH"
1757. 2428
/gene="STY3365"
1757. 2428
/gene="STY3365"
/note="Orthologue of E. coli ygiB (YGI_B_ECOLI); Fasta hit
to YGI_B_ECOLI (234 aa), 97% identity in 223 aa overlap.
Contains a possible N-terminal signal sequence."
/codon_start=1

Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
 Db 15313 GCGTCTCTGGAGGAAAGCCCGCTGCTGGCTGAGCGGTATTGATGAGCGCGCGC 15254
 Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
 Db 15253 GCGCTGAGCAAAAGCGTGGCGGAGGTCAATGTGAAGTGCAGCTTCGTGTCGCCGACG 15194
 Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
 Db 15193 CATCCGAGGATTACCAACTGCGCGTACTATCAGCTAATCAGCAGCTCATTCGCTTGAT 15134
 Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
 Db 15133 TTTCAAGAAAGCTTTGAGGGCGTGAGCCCGCAGCGCTTGCGATGAGCTATCAACGAGCG 15074
 Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
 Db 15073 CTGGGATCGATCGCGCGCTGTATTGTCGATATTATGCCAAGGCGCTGTGACGAGCGTG 15014
 Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
 Db 15013 CAGACTATGATTTCCCTAGCCGCGCAGCGCGCTGCGGTGCTCATCGATCCGAAAGGA 14954
 Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
 Db 14953 ACGGATTTGAACGTTACCGCGCGCCACGCTGCTGAGCGCAACCTTCGTGAATTCGAG 14894
 Qy 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
 Db 14893 GCGGTTGCGGGAATGTAAAGCGAAGACGAACTGTTGAACGCGCATGAAGCTCAT 14834
 Qy 221 AlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
 Db 14833 GCCGATTACGACCTTCGCGCTGTTGTCACGCTTCGCAACAGGGAATGACGCTGCTG 14774
 Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
 Db 14773 CAACCGAATAAGCGCGCTACATATGCCGACGCGCGCAGGAGTTTATGATTTACC 14714
 Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
 Db 14713 GGTGCGGCGGATACGCTGATCGCGCTGCTGCGCGCGAGCTGGCGGCGGAATACCGTG 14654
 Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
 Db 14653 GAGGAGCGGTATTATTCGCCAATGCGCGCGCGGTAGTGGTAGGTAACCTCGGAGC 14594
 Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
 Db 14593 TCAACGGTTTCCCTATTGAGCTGGAAACCGCAGTGGCGGACGCGCGGATACCGGCTTC 14534
 Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaArgLysArgGlyGlu 340
 Db 14533 GCGCTCATACCGAAGAGGAGCTTGAGACAGCGCTGCGCAGCGCGGTGAAGCGTGGCGAG 14474
 Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
 Db 14473 AAATGGTTCATGACCAACGCGGCTTTTCGATATTCGACGCGGCGCCACGCTCTTATCTG 14414
 Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
 Db 14413 CCGAACGCGCGCAACTGGGCGCAGCGCTGATTGTCGGTCAATAGTACGCGCTCGACT 14354
 Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
 Db 14353 AAACGCTCTGAAGGCGAAAGCGCTCCGGTTAATCCGCTCGCAACAGCGTATGATCGTGTG 14294
 Qy 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420
 Db 14293 GCGCGCTGGAGTCGCTGAGTGGGTGCTCTTTTGAAGAGGATACGCGCAACGACTG 14234
 Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440

Db 14233 ATTCCCGGTATTCTGCGGATCTGCTGTAAGGCGCGGACTATAGCCGGAAGATC 14174
 Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 Db 14173 GCGGCGAGCAAGAGGTCTGGGCAACGCGCGCAAGTCATGGTCTGAACCTCGAAGAT 14114
 Qy 461 GlyCysSerThrAsnIleIleLysLysIleGlnGlnAspLysLys 476
 Db 14113 GGTGTTCCAGGACCAATATCATCAAAAGATCCAGACCGGAGCGGAG 14066
 RESULT 10
 AF163661
 LOCUS
 DEFINITION Salmonella enterica RfaE (rfaE) gene, complete cds.
 ACCESSION AF163661
 VERSION AF163661.1 GI:5734093
 KEYWORDS
 SOURCE
 ORGANISM
 Salmonella enterica.
 Salmonella enterica
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Salmonella.
 REFERENCE
 1 (bases 1 to 1476)
 AUTHORS Valvano,M.A., Marolda,C.L., Bittner,M., Glaskin-Clay,M., Simon,T.L. and Klens,J.D.
 TITLE The rfaE gene from Escherichia coli encodes a bifunctional protein involved in biosynthesis of the lipopolysaccharide core precursor
 JOURNAL J. Bacteriol. 182 (2), 488-497 (2000)
 MEDLINE 20096694
 PUBMED 10629197
 REFERENCE
 2 (bases 1 to 1476)
 AUTHORS Valvano,M.A.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-1999) Microbiology and Immunology, University of Western Ontario, DSB3003, London, Ontario N6A5C1, Canada
 FEATURES
 source
 1. .1476
 /organism="Salmonella enterica"
 /strain="SL1027"
 /db_xref="taxon:28901"
 1. .1434
 /gene="rfaE"
 1. .1434
 /gene="rfaE"
 /function="bifunctional enzyme involved in the synthesis of ADP-glyceromannoheptose"
 /note="putative ADP transferase"
 /codon_start=1
 /transl_table=11
 /product="RfaE"
 /protein_id="AAD49846.1"
 /db_xref="GI:5734094"
 /translation="MRVNLPAFERAGVVMVDMLDRYWGPTCRISPEAPVPVVKVN
 TVBERPGCAANVAMNIALSGANARLVGLTGIDDAARALSKTLEAVNYKCDPVSYPHTP
 TITKRLVLSRNOQLIRUDFEGFEGVPOPLHERINOALSIGALVLSDAKALTSV
 QTMISLARQAGVPLIDPKGTDERYRGATILLTNLSEFEAVACKCKSEDELVERGK
 LIADYLSALLVTRSEGMILLQTPKNKAPLHMPTOAQEYDVTGAGDVIQVLAATLAA
 LNTLEACYFANAAGVYVVGKLTSTVPIELANAVRPPDTGFGVMTKEELRQAVP
 ARKRGKVMATNGVFDILHAGHSYLANARKLDRLLIVAVNSDASTKRLKESRPVNP
 LEORMIVLGALVESVDVWVSPEEDTPQRLIAGILPDLLVKGGDKPEETAGSEEWANG
 GEVNLNFDGCSPTTIKIQTSEK"
 1. .954
 /gene="rfaE"
 /note="similar to ribokinase family; Region: domain I"
 1030. .1431
 /gene="rfaE"
 /note="similar to cytidyltransferase family; Region: domain II"
 319 a 372 c 481 g 303 t 1 others
 BASE COUNT
 ORIGIN
 Alignment Scores:


```
/protein_id="AAD49847.1"
/db_xref="GI:5734096"
/translation="MRVNLPAFERAGVMVGDMLDRIYWGPTCRISPEAPVPVKKVN
TVEERPGGAANVAMNIALSGANARLVGLTGIDDAARALSKTLAEVNVKDFSVPTHP
TITKRLSRNOOLRLDFEGFEGVDFQPLHERINQALGSLGALVYKDALGTSV
QTMISLARGVLLIDPKGDTDFERYRGATLLTPNLSEFAVAGKCKSEDELVERGMK
LIADYDLALIVTSEQEMTLIQPNKAPLHPMTQAEYVDVTGAGTIVGLAAGNTL
EACVFANAAAGVVGKLGTSVSIENLNAVGRADTGFVMTVEELRQAVASARKR
GEKVMTNGVDILHAGHVSILANARKLGDRLIVAVNSDASTKRLKGRSPVNPLEOR
MVLGALSDVWVYSFEEDTQRLIAGLIPDLLYKGGDYKPEEIASGSEVWANGGEV
VUNFEDGGSTNIIKKIQTESEK"
variation
707
/gene="rfae"
/note="glycine to glutamic acid substitution"
/replacement="g"
818-819
variation
/gene="rfae"
/note="deletion of four amino acids"
319 a 366 c 475 g 304 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 2,73e-128 Length: 1464
Score: 2225.00 Matches: 439
Percent Similarity: 96.43% Conservative: 20
Best Local Similarity: 92.23% Mismatches: 13
Query Match: 92.98% Indels: 4
DB: 1 Gaps: 1
US-09-912-020-325 (1-477) x AF163662 (1-1464)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAGAGTAAATCGCACGGTTGAAACGTTGACGGCGTATGTTGGTGGTGAATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTATTGGTATGCGCCCACTTGCCTGATTTCACCGAAGCGCGGTGCCCGTG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGGTAATACCGTTGAGGAACCCCGCGCGCGCGAAGCGGCGGATGAACATT 180
QY 61 AlaSerLeuGlyAlaAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCCTCTCTGGGACGAAACCCGCTCTGCTCGGCTGACGGGTATGATGACGCCGCCGCG 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCCTCAGCAAAACGCTGGCGGAGGTCAATGTGAAGTGGCACTTCGTTTCTGTGCCGACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACGATTACCAAACTGGCGGTACTATCACGTAATCAGCAGCTCATTCGCTTGTAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAGGCTTTGAGGCGGTGGACCCGCGCGCGCTTCATGAGCGGTATCAACAGGCG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGGGATCGATCGCGCGTGTATTGTCGGATTATGCCAAAGCGCTCTGACACGCGTG 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CACAGATATGATTTCCTGGCGGCCGAGCGCGCTCGCGGTGCTCATCGATCCGAAAGGA 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACGGATTTGAAGCTTACCGCGCGCGCGCTGCTGACGCCCAACCTTCTGAAATTTGAG 600
QY 201 AlaValValcylLysCysLysThrGluGluGluIleValcylArgGlyMetLysLeuIle 220
Db 201 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
Db 601 GCGGTGCGCGGGGAAATGTAAGACGAGAACGAGTGTGTTCAACGCGGCATGAACACTCAT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCTGATTACGACTTTCCCGCGCTGTTGGTCCACCGTTCGACACGAAATAGCGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGAATAAAGCGCGCTACATATGCGCGACGCGCAGCAAGCTTTATGATGTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGCGATACGCTGATCGCGGTG-----CTGGCGCGGAAATACACCTG 828
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 829 GAAGAGCGGTATTATTTCGCCAATGCGCGCGGTAGTGGTAGTAACTTCGGGACG 888
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 889 TCAACGCTTCCCTATTGAGCTGGAACAGCAGTCCGCGGACGCGGATACCGGCTTC 948
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 949 GCGGTTATGACCGACAGAGGATGACACAGGCGTCCGACGCGCGTAAAGCGTGGCGAG 1008
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1009 AAAGTGGTTCATGACCAACGCGGTTTCGATATTCTCAGCGCGGCGCATGCTCTTATCTG 1068
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1069 GCGAAGCGCGCAAACTGGCGACCGCTGATGTTGCGGTCAATAGTACCGCTCGACT 1128
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1129 AAACGCTGAAGGCGGAAAGCGCTCGGTTAATCCGCTCGAACAGCGTATGATCGTCTA 1188
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1189 GCGCGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1248
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyLysGlyAspTyrLysProGluGluIle 440
Db 1249 ATTGCGGCTATTCTGCGGATCTGCTGTAAGGCGCGGCTATTAAGCGGGAAGATC 1308
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1309 GCGGGCAGCGAAGAGGTCTGGGCAACGCGGTGAAGTCAATGCTGCTGAACCTCGAAGAT 1368
QY 461 GlyCysSerThrThrAsnIleLysLysIleGlnGlnAspLysLys 476
Db 1369 GGTGTTCCACGACCAATATCATCAAAAGATCCAGACCGAGAGCGAG 1416
RESULT 12
LOCUS AE013955/1 10641 bp DNA linear BCT 26-JUL-2002
DEFINITION Yersinia pestis KIM section 355 of 415 of the complete genome.
ACCESSION AE013955
VERSION AE013955.1 GI:21960480
KEYWORDS Yersinia pestis KIM.
SOURCE Yersinia pestis KIM.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.
REFERENCE 1 (bases 1 to 10641)
AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.
TITLE Genome Sequence of Yersinia pestis KIM
JOURNAL J. Bacteriol. 184 (16), 4601-4611 (2002)
```

PUBMED	12142430		
REFERENCE	2 (bases 1 to 10641)		
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		
FEATURES	Location/Qualifiers		
Source	1..10641		
	/organism="Yersinia pestis KIM"		
	/strain="KIM"		
	/db_xref="taxon:187410"		
gene	complement(234..1016)		
	/gene="y3519"		
CDS	complement(234..1016)		
	/gene="y3519"		
	/note="residues 1 to 260 of 260 are 70.99 pct identical to residues 10 to 271 of 271 from E. coli K12 : B3039; residues 1 to 260 of 260 are 72.13 pct identical to residues 10 to 271 of 271 from GenPept : >g AAG58178.1 AE005533.11 (AE005533) orf, hypothetical protein [Escherichia coli O157:H7 EDL933]"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAM87067.1"		
	/db_xref="GI:21960481"		
	/translation="MNTFRPALFLGHGSPMNVLENSYTAQWRTIGETLPRPKMILA ISAHYTRGATVTAHEHPTIHDFFGFPQALFDIQPAPGSELAHIOQLLAPTPVR ADTGEWLDGSGWLVLMKYPEADIPVVLQSIDGTQPAAYHYELGOKLAALREQGVMI VASGVNHLRLGKWOCESSPYPAWESFNQVRDNLSYOGDDHPLVNEHQEGEALSN PSEHYLLPLVYLGSDGKEPSIPTDGVEMGSLMSLVQIG"		
gene	complement(1441..2094)		
	/gene="ribB"		
	/note="y3520"		
CDS	complement(1441..2094)		
	/gene="ribB"		
	/function="enzyme; biosynthesis of cofactors, carriers: Riboflavin"		
	/note="residues 1 to 217 of 217 are 77.41 pct identical to residues 1 to 217 of 217 from E. coli K12 : B3041; residues 1 to 217 of 217 are 78.34 pct identical to residues 1 to 217 of 217 from GenPept : >g AAL22069.1 (AE008847) 3,4 dihydroxy-2-butanone-4-phosphate synthase [Salmonella typhimurium L72]"		
	/codon_start=1		
	/transl_table=11		
	/product="3,4 dihydroxy-2-butanone-4-phosphate synthase"		
	/protein_id="AAM87068.1"		
	/db_xref="GI:21960482"		
	/translation="MNQTLISDFGTPVERVERAIDALNRGRGMVLLDDESRENGDMV FAAPAMLEQALTIRGSGICITDERRQQLDLPMMVTHNSSQFQTAFTVTEAA EGVTTGVSADRLTTIRKATADNAKPADLNRPCHVFLRQGVLSRRHGTEASIDL ATLAGYKPAVGLCELTNDGCSMAHAPEVIAFAKLHDMPPVTIDDLAAAYLQSRKKAS"		
gene	2145..2627		
	/gene="y3521"		
CDS	2145..2627		
	/gene="y3521"		
	/function="unknown"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical"		
	/protein_id="AAM87069.1"		
	/db_xref="GI:21960483"		
	/translation="MRSKRPKIKSRNNASGPYPLFSKRQRCLWFGFPQIAYSCKHLRI FVMTNAPHLIETRICISPADVTLSHPDYNRRPNYTGSDALYLALIDKENNSARGLL HHRDVIYATQTAGGDFHPALISPTWLTALLIHLGNYQNAELNASRGLYNQHRALY		
gene	2663..2944		
	/gene="y3522"		

PQLLLMKGKLGSELNFSDDIDLEIFAYPENGOTQGRRELNDNAQFTRILGRLIKA
 LDQRTIDGFVYRDMRPFDSGLPSVLFAALDQYQEOGRDWERVAMVKARUMGGA
 DDVYSLOKRLMRFVRRXIDFSVIOSLRNMKGMIAREVRRLKDNKIKLAGGIRE
 LEFTOVQLIRGGRERLQERALLPILOVAELGLLPEQOVADLSYFLRLRENL
 LOATADQOTLPNDPQNLKAWMGTAADWAAAMSTALENHQAVRVFDDLIGDET
 LGEDPSHGLYSUQDVLEESDAPLTPHLEAAROLLATISGFRRHVDKRTIGPR
 GREYLDMPFLFAEVCPRPDANVALSRLILLLSIVRTTYLELLVEYHAALKHVIR
 LCSAPMVASQLARYPLLDDELDPQSLYQPLAPSAVRDELROYLLRPVDEDEBQOLE
 ALROFKAOLRIAGDITREALPVMSVDHLTYLAEALIDAVIOQANOMVARYGQPS
 HLOQSEGRFAGVIGYKGLGWELGYSSELDLVLFDLPDMDTGDRLDGRLQRYLRL
 AQRIMHLSFTSSGILYEVDARLRPSGEAGMLVSTIEAFDYORNEAWTWEHQAIR
 ARIVYSPKLUQFDALRQOILKRRHEDPOLQOQVREMRKMRNLHGLSKQRDIPDIKA
 DAGGITDIEFIAQYLVLYRYAASEPLRTFWSNDVRFESMAHYDIMSPEEAALTRAVV
 TMRDEIHLALQEQQSSKVAADSFIAEREQVAAASHKWLAAANDANVS"
 complement(9147..9233)
 /gene="y3526"
 /complement(9147..9233)
 /gene="y3526"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /product="hypothetical"
 /protein_id="AA087074.1"
 /db_xref="GI:21960488"
 /translation="MARQGHWRDTRARKEALSQAFLWNGAVR"
 complement(9285..10466)
 /gene="y3527"
 complement(9285..10466)
 /gene="y3527"
 /note="residues 5 to 392 of 393 are 51.54 pct identical to
 residues 1 to 386 of 433 from E. coli K12 : B3054;
 residues 5 to 392 of 393 are 52.31 pct identical to
 residues 1 to 386 of 433 from GenPept : >emb|CAD07727.1|
 (A1627278) conserved hypothetical protein [Salmonella
 enterica subsp. enterica serovar Typhi]"
 /codon_start=1
 /transl_table=11

gene

CDS

gene

CDS

Alignment Scores:

Pred. No.:	4.28e-118	Length:	10641
Score:	2076.00	Matches:	409
Percent Similarity:	92.84%	Conservative:	32
Best Local Similarity:	86.11%	Mismatches:	34
Query Match:	86.75%	Indels:	0
DB:	1	Gaps:	0

US-09-912-020-325 (1-477) x AE013955 (1-10641)

Qy	1	MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet	20
Db	5986	ATGAAAGTCACGCTGCCTGATTTTCGCCGCTGCTGGTGATTTAGTCGGTGCATGTCATG	5927
Qy	21	LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal	40
Db	5926	TTAGACCCGCTACTGTGACGGGCCAACCTGCGCGGATTTCCACAGAACGACCTGTGCGCTGTA	5867
Qy	41	ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle	60
Db	5866	GTGAAAGTCGATACCATCGAAGAGCGCCCTGGTGGTCAGCAACGTTGCGATGAATATT	5807
Qy	61	AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg	80
Db	5806	GCCTCCCTTGGCGGGTTGCGCGCTTAGTCGGGTTGACCGCATTTGACGATGCGCGCGGT	5747
Qy	81	AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr	100
Db	5746	GGCTGTGATCTCGAAGCTGAGCGAAGTTCGGGTGCGGATTTGCTCGGTACCCACT	5687
Qy	101	HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp	120
Db	5686	CACCGGACCATACCAATACGGGTACTTTCCGCTACACGACGAGTTAATCCGCTGGAT	5627
Qy	121	PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla	140

Db	5626	TTTGAAGAGGGCTTTGACGGTGTTCATCCACGCCGATATTCAGCGCATTCACACTGGCG	5567
Qy	141	LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal	160
Db	5566	TTGCCACAGATTGGCGCGCTGGTGTCCGATTATGCAAGGGGCGATTGAACAGCGTA	5507
Qy	161	GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly	180
Db	5506	CAGCGCATGATCCAACTGGCAGCTAAAGCAATGTCCTCGGTTTGTATTCATCCGAAGGC	5447
Qy	181	ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu	200
Db	5446	AGTGATTTGAGCGCTACCGTGGTCCACATTTGTGACGCCGCAATTTGCTGAATTTGAA	5387
Qy	201	AlaValValGlyLysCysLysThrGluGluLleValGluArgGlyNetLysLeuIle	220
Db	5386	GCAGTGGTAGGCGCTGTAATAATGAAGAAGATTGGTTAACCCGCGGTATGCAACTGGTG	5327
Qy	221	AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu	240
Db	5326	GCAGATTTTCGAACITTCAGCATTTGCTGGTGACCCGTTCTGAACAAGGCATGACCTTGTG	5267
Qy	241	GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr	260
Db	5266	CAACTGGGTAAACGCCCTACATTTGCCGACTCAGGCAAAAGAAAGTTGTTGATGTGACC	5207
Qy	261	GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu	280
Db	5206	GGTGTGCGGATACAGTTTATTTGGCGTGTGCTGGCGCGCTGGCTGCAGGTAAACAGCTG	5147
Qy	281	GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr	300
Db	5146	GAAGAGTCATGTTCTCGCTAAATGCTGCCGCGGTAGTGGTGGTAAATTTGGGTACC	5087
Qy	301	SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe	320
Db	5086	TCGACCGTTTCTCCGATTGAGTTGGAAAAACGCGATTCCTGCGCGTGCAGAAAACCGGTTT	5027
Qy	321	GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAla	340
Db	5026	GGGTGTGATGGATGACGACCACTGAAATAGCGGTGGCGAGGCTCGCCAGGTGGGGAG	4967
Qy	341	LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu	360
Db	4966	AAGGTGGTCATGACCAATGGCATTTTTCACATCTCTCCATGCGAGGCCATGTTCTTATCTG	4907
Qy	361	AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr	380
Db	4906	GCCAATGCCGCTAAGCTGGGGATGCGCTGATTTGTGGCGGTGAACAGCGCATGCTTCCACC	4847
Qy	381	LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu	400
Db	4846	AAGCGGTGAAAGGTGAAGCGCGGTTAATCCTTAGAACACACGATGATGTTGCTCTTG	4787
Qy	401	GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu	420
Db	4786	GGTGGTTTGAAGCGGTGGATTGGTCTGCTGCTTTTGAAGAAGATACCCACAGCGTTTA	4727
Qy	421	IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle	440
Db	4726	ATTGCTGATATTTTCCCGCATCTGTAGTCAAGCGCGCGATTATTAAGCCCATGAAAT	4667
Qy	441	AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp	460
Db	4666	GCAGTACGAAAGAGTCTGGCGCCCGCTGGCGAGGTTAAGGTACTGAATTTTGAAGAT	4607
Qy	461	GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLys	475
Db	4606	GGTGTCTCTACCAACCAATATTATTCAGTCGATCAAAACACGCCGCG	4562

 RESULT 13
 AJ414144

LOCUS AJ414144 208050 bp DNA linear BCT 06-JUN-2002
DEFINITION Yersinia pestis strain C092 complete genome; segment 4/20.
ACCESSION AJ414144 AL590842
VERSION AJ414144.1 GI:15978734
KEYWORDS Yersinia pestis.
SOURCE Yersinia pestis
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.
REFERENCE 1 (bases 1 to 208050)
AUTHORS Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Fellwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karylshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrrell, B.G.
TITLE Genome sequence of Yersinia pestis, the causative agent of plague
JOURNAL Nature 413 (6855), 523-527 (2001)
MEDLINE 21470413
PUBMED 11586360
REFERENCE 2 (bases 1 to 208050)
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT Notes:
 Details of Y. pestis sequencing, at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES Location/Qualifiers
 1..208050
 /organism="Yersinia pestis"
 /strain="C092"
 /db_xref="taxon:632"
 /note="biovar: Orientalis"
 complement(92..307)
 /gene="YPO0645"
 /note="rpsu"
 complement(92..307)
 /gene="YPO0645"
 /note="Identical to Escherichia coli and Salmonella typhimurium 30S ribosomal protein S21 Rpsu SW:RS21_ECOLI (P02379) (70 aa) fasta scores: E(): 2.3e-27, 100.0% id in 70 aa, and to Buchnera aphidicola (subsp Acyrthosiphon pisum) 30S ribosomal protein S21 Rpsu TR:BA012780 (EMBL:AP001118) (71 aa) fasta scores: E(): 1e-25, 90.1% id in 71 aa"
 /codon_start=1
 /transl_table=11
 /product="30S ribosomal protein S21"
 /protein_id="CAC89499.1"
 /db_xref="GI:15978735"
 /db_xref="SWISS-PROT:P02379"
 /translation="MPVIKVRENEPFDVALRRFRKSCERKAGVLAERVRRFEYKPTTE RKRKASAVKRIAKKLARENARRTRYL"
 complement(143..304)
 /gene="YPO0645"
misc_feature
 /note="Fram match to entry PF01165 Ribosomal_S21, Ribosomal protein S21, score 127.90, E-value 1.9e-34"
 complement(233..271)
 /gene="YPO0645"
misc_feature
 /note="PS01181 Ribosomal protein S21 signature."
 713..1726
 /gene="YPO0646"
 /note="gcp"
 713..1726
 /gene="YPO0646"
 /note="Similar to Haemophilus ducreyi sialylglycoprotease Gcp TR:Q9L7A5 (EMBL:AF224466) (348 aa) fasta scores: E(): 0, 73.4% id in 342 aa, and to Escherichia coli probable O-sialoglycoprotein endopeptidase Gcp SW:GCP_ECOLI

(P05852) (337 aa) fasta scores: E(): 0, 87.8% id in 335 aa"
 /codon_start=1
 /transl_table=11
 /product="putative glycoprotease"
 /protein_id="CAC89500.1"
 /db_xref="GI:15978736"
 /db_xref="SPTREMBL:Q8Z168"
 /translation="MRVLGIETSCDETGIADVDDKAGLANOLYSQVKLHADYGGVVP ELASRDHVRKTYPLIOALKKEANLSAKDIDAVYTAGCLVGLVGLATIGRALAFW GVPAPVHHMEGLHAPLEENAPPEFVALVYSGHTQLISVTGIGELLGESVDD AAGEAFDTAKLLGLDYGPGMLSRMAQGTGVRFTFPMTDPRGLDFFSGLKTFE ANTIRANGDDQTRADIRAFEDAVVDTLAIKRALDQTFGRKLVIAAGVANSQTLR LKLDMMQKRGGEVYARPEFCTDNGMIAYAGMVRLSNLSNLSVSVRWRPLSEL PKV"
 713..1654
 /gene="YPO0646"
 /note="Pfam match to entry PF00814 Peptidase_M22, Glycoprotease family, score 649.20, E-value 2.2e-191"
 998..1060
 /gene="YPO0646"
 /note="PS01016 Glycoprotease family signature."
 complement(1972..2622)
 /gene="YPO0647"
 complement(1972..2622)
 /gene="YPO0647"
 /note="Similar to Escherichia coli hypothetical protein YgiH SW:YGIH_ECOLI (P31056) (205 aa) fasta scores: E(): 0, 82.8% id in 204 aa, and to Haemophilus influenzae hypothetical protein HI0266 SW:YGIH_HAEIN (P44603) (199 aa) fasta scores: E(): 0, 67.3% id in 199 aa"
 /codon_start=1
 /transl_table=11
 /product="putative membrane protein"
 /protein_id="CAC89501.1"
 /db_xref="GI:15978737"
 /db_xref="SPTREMBL:Q8Z167"
 /translation="NSAIALGMIIFAYLCGSISSAILVCRVARLPDPRTHSGNPGAT NVLRIGRTAAVAVLFDILKGMIPWAIYLLHPIPLYLGLTAACLGHIYVPFHF KGGKVATFAGAIPIGWDLTGIMTGTWLLTVLLSGYSIGAIYALPFFVWFKP QFTFPVAMLSCLILMRHNDIQRLWRGREGKIWDKLRKKQTPAEAAELEKEK"
 complement(2149..2214)
 /gene="YPO0647"
 /note="one of 5 probable transmembrane helices predicted for YPO0647 by TMHMM2.0"
 complement(2230..2295)
 /gene="YPO0647"
 /note="one of 5 probable transmembrane helices predicted for YPO0647 by TMHMM2.0"
 complement(2317..2382)
 /gene="YPO0647"
 /note="one of 5 probable transmembrane helices predicted for YPO0647 by TMHMM2.0"
 complement(2395..2460)
 /gene="YPO0647"
 /note="one of 5 probable transmembrane helices predicted for YPO0647 by TMHMM2.0"
 complement(2545..2610)
 /gene="YPO0647"
 /note="one of 5 probable transmembrane helices predicted for YPO0647 by TMHMM2.0"
 2730..3089
 /gene="YPO0648"
 /note="folB"
 2730..3089
 /gene="YPO0648"
 /EC_number="4.1.2.25"
 /note="Similar to Escherichia coli dihydroneopterin aldolase FolB SW:FOLB_ECOLI (P31055) (122 aa) fasta scores: E(): 0, 84.7% id in 118 aa, and to Vibrio cholerae putative dihydroneopterin aldolase VC0524 TR:Q9KJ06 (EMBL:AE004138) (129 aa) fasta scores: E(): 2.9e-30, 67.2% id in 116 aa"

CDS	4118..4270 /gene="H11524" /note="hypothetical protein; identified by GeneMark; putative" /codon_start=1 /transl_table=11 /product="H. influenzae predicted coding region H11524" /protein_id="AAC23181.1" /db_xref="GI:1574374" /translation="MIPTEGIHLLIKGSSSANIPCELSKSLSPVTNGKTIKSN KVSND" complement(4508..5245) /gene="H11525" complement(4508..5245) /gene="H11525" /note="similar to GB:AE000666 percent identity: 32.51; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="molybdate-binding periplasmic protein, putative" /protein_id="AAC23171.1" /db_xref="GI:1574366" /translation="MKLVAVTSMILTFVSQAADLYIAGAGLKEPVEKIIHQYEQ E TGNVTVIGSGQLARINTVKSDDLFGAGEDYVTKLQTNNDVNNIGTVLHVPM AIRKDTSGIDSFALAESSLRIGDGSKAMALGKAEMFELSGYQQLNDKIVKA ATVQQLMLYLLNGDVAADVGRSGAWKVRDVELLPSPKGTPEKVTIGLFFSKYKPK EAQQLDFPKSPQGVKVFTEGFLPAK" complement(5432..6862) /gene="H11526" complement(5432..6862) /gene="H11526" /note="similar to PID:882153 percent identity: 97.38; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="Adp-heptose synthase (rfae)" /protein_id="AAC23172.1" /db_xref="GI:1574367" /translation="MAOYSAEFKQAKVLVLGVMDLDRYWFQGNATRISPEAPVVRVQ ENERGAAGAAVANNIASLNPVOLMGLIGDGTCSALSLLEKQIDCNVALETHP TITKLRLSRHQQLRLDFEDFNNVCKDLAKLESVKNYKALILSDYKGLTKDV QKMQIARKANVLLIDPKGDFYRGATLLTPNMSEFEAVVKGKCTEBEIEKGLK LSDIETALLTVSRKGMTLLRNPQEPYHLPTVAKEFVDVTGADTVISYLATALAD GRSFEESYLANAAGIVGKLGSTVSTVELENAIHARPETGFMSEAELEKDAVAQ AKARGEITVMNGCFDILHPGHSYLENARKLGRILVAVNSDVSVKRLKCESRPIN LENMAYVLASVDVMDLYPFTEDTPRIGELPLDLLVKGDDYRPEETAGSKEVWANG GDVRLNFENGCSSTNVIEKILLND" 6947..7894 /gene="H11527" 6947..7894 /gene="H11527" /note="similar to SP:P24187 GB:X61000 PID:48957 GB:U00096 PID:1651518 percent identity: 55.88; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="lipid A biosynthesis lauroyl acyltransferase (htxB)" /protein_id="AAC23173.1" /db_xref="GI:1574368" /translation="MERTMKNELKQFQHPFLAPKYLFWLGVAIWRSILCLPYPILR HIGHGFWSLHAKVGERRAARINLELCPDMPENREVILOENLRSVGMAIETG MAWFSQSRIRKKSVEGLHLYKENQKDGIVLVGVHETLELGARILGLHHPGICVVR PNDPLDLWLOTQELRSNKMDLRKGLRMKALRHEETIHWAPDHDYGRKNVFP FAFVPCVTITGSIYLLKSSQNSKVIPIFAPLRNKDGSITVSIAPVDFDLDQDETAI AARNQIVEIKEDITITMMLHRRFTRPDEKTPSLYD" 7968..9866 /gene="H11528" 7968..9866 /gene="H11528" /note="similar to SP:P20083 GB:L22026 GB:M58409 PID:147108	Pred. No.: 1.43e-95 Length: 14465 Score: 1712.00 Matches: 331 Percent Similarity: 83.48% Conservative: 58 Best Local Similarity: 71.03% Mismatches: 77 Query Match: 71.54% Indels: 0 DB: 1 Gaps: 0 US-09-912-020-325 (1-477) x U32828 (1-14465) Qy 7 GluPheGluAraGlaGlyValMetValValGlyAspValMetLeuAspArgTyrTrpTyr 26 Db 6844 GAATTTAAGCAAGCAAGTACTGCTATTAGCGGAGTGTCTGATCTGATTGGTTC 6785 Qy 27 GlyProThrSerArgIleSerProGluAlaProValProValValValValAsnThrIle 46 Db 6784 GCGGCAACCAACCGTATTTCCACGAGGACCAAGTACAGTGGTTCGTACAGAAAT 6725 Qy 47 GluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn 66 Db 6724 GAAGAACGCGGGTGTGCGCAAAATGTGGCATGAATATTGCTTCACTCAATGTACCC 6665 Qy 67 AlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLeu 86 Db 6664 GTTCAGTTAATGGGATTTGATTGGCAAGATGAACCTGCTGCACATTTCCCTCTTATTA 6605 Qy 87 AlaAspValAsnValIlyCysAspPheValSerValProThrHisProThrIleThrLys 106 Db 6604 GAAAAACAAAAAATGATTGTAATTTTGTGCATTAGAACCCCATCAACATTACTATA 6545 Qy 107 LeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyPheGlu 126 Db 6544 TTACGTATTTATCTGTCATCAACAGCTGCTCCGCTTGATTTTGAAGAAGATTTCAAT 6485 Qy 127 GlyValAspProGluProLeuHisGluArgIleAsnGlnAlaLeuSerIleGlyAla 146 Db 6484 AATGTAGATTGCAAGGATTTATTAGCGAAGTTAGAAAGTGGCGGTGAAAATTTACGGTGT 6425 Qy 147 LeuValLeuSerAspTyrAlaLysGlyAlaLeuAspValGlnGlnMetIleGlnLeu 166 Db 6424 TTGATCTTCTGATTACGCCAAGCCAGCTTAAGATGTTTACAGAAATGATTCAAAT 6365 Qy 167 AlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArgTyr 186 Db 6364 GCACGCAAGCGAATGTGCTGCTGTTGATCGATCCAAAGGAAGCACTGATTTGAACGTTAT 6305 Qy 187 ArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValGlyLysCys 206 Db 6304 CGTGGGCTTACATTATTGACACCAATATGCTGAATTTGAAGCTCTTCTGGGTAAATGC 6245 Qy 207 LysThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSer 226 Db 6244 AATACGGAAGACAGATTTATTGAGAAAGGTTTAAATTAATTTCTGATATTCAATTAACC 6185 Qy 227 AlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaPro 246 Db 6184 GCACTTTTGTCGACGCTTCTGAAAAAGCATGACATTTATTACGCCCAATCAAGAGCCT 6125 Qy 247 LeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrVal 266 Db 6124 TATCATTTTCCCAACTGTTTCCAAAGAAAGTGTGTGATGTCAGCGGAGCCGGTGACACTGTC 6065 Qy 267 IleGlyValLeuAlaAlaThrLeuAlaLaGlyAsnSerLeuGluAlaCysPhePhe 286 Db 6064 ATTACGATTATAGCAACCCCATTTAGCAGATGCGCTTCTTTTGAAGATCTTTGTACCTA 6005 Qy 287 AlaAsnAlaAlaGlyValValValGlyLysLeuGlyThrSerThrValSerProIle 306 Db 6004 GCCAATGTTGCCGAGGAATTTCTGTGGGAAATTTGGGGACTTCAACGCTTCGACCGTG 5945 Qy 307 GluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGlu 326 Db 5944 GAACTTGAAATGCGATTTCATGCTGCTGCTGAACTGGATTTGGCATTATGAGTGAAGCA 5885
-----	---	--

Db 750 GAATTAGAAAATGCAATTCACGGCTGTTCTAATACCGGATTGGTATTATGAGCGAAGCC 691
Qy 327 GluLeuLysLeuAlaValAlaAlaArgLysArgGlyGluLysValValMetThrAsn 346
Db 690 GAATGAAACCGGTAGTGGCCCAAGCGGCGGTGGTGAAGAAATGTGATGACTAAT 631
Qy 347 GlyValPheAspIleLeuHisAlaGlyHisValSertyrLeuAlaAsnAlaArgLysLeu 366
Db 630 GGCTGTTTGTGACATTCACCCGGGACATGTGCTTACCTCGAAAACGCGCAATTTG 571
Qy 367 GlyAspArgLeuIleValAlaValAsnSerAspAlaSerThrLysArgLeuLysGlyAsp 386
Db 570 GCGCATCGCTTAATGTGCGCAGTGAACAGCGATGACTAGTCAACCGCTAAAAAGGTGAA 511
Qy 387 SerArgProValAsnProLeuGluGlnArgMetIleValLeuGlyAlaLeuGluAlaVal 406
Db 510 ACCCGCCCAATATGACCTTCAATCTCGTATGGCAGTATTGGCAGGCTTATCTTCAGTC 451
Qy 407 AspTrpValValSerPheGluGluAspThrProGlnArgLeuIleAlaGlyIleLeuPro 426
Db 450 GATTGGTTAGTGGCGTTCATGAAGATACACCAACGTTTAATTGCTGAAGTATTACCT 391
Qy 427 AspLeuLeuValLysGlyAspTyrLysProGluGluIleAlaGlySerLysGluVal 446
Db 390 GATTATTGGTGAAGGTGGCGATTACAACCAAGAGACATCGCCGGTAGCCAAAGATT 331
Qy 447 TrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAspGlyCysSerThrThrAsn 466
Db 330 TGGCGGAGTGGTGTGATGATAAAGTATTAAACCTTTGAAAATGGTTGTTCAACCGATTAAT 271
Qy 467 IleIleLysLysIleGlnGlnAspLys 475
Db 270 GTGATCAGTAAGATTCAGCAGCTTAAAA 244

Search completed: November 26, 2002, 22:04:32
Job time : 3572 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2002, 21:38:24 ; Search time 82 Seconds
(without alignments)
775.129 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLPEFERAGVGVGVDM.....FEDGCTTIKIQDKKG 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	2393	100.0	477	21	AA15968	E. coli proliferat
2	2393	100.0	477	22	AAU34708	E. coli cellular p
3	2393	100.0	477	22	AAG98421	Escherichia coli p
4	2265	94.7	477	22	AAU38166	Salmonella typhi c
5	1350.5	56.4	474	22	AAU36486	Pseudomonas aerugi
6	745.5	31.2	323	23	AAU72936	Neisseria meningit
7	734.5	30.7	424	18	AAW20763	H. pylori cytoplas
8	610.5	25.5	498	21	AAW07577	Protein encoded by
9	197	8.2	308	22	AAU36261	Pseudomonas aerugi
10	191.5	8.0	305	22	AAU35299	Enterococcus faeca

ALIGNMENTS

RESULT 1

AA15968

ID AA15968 standard; Protein: 477 AA.

XX AA15968;

XX 05-OCT-2000 (first entry)

DT 05-OCT-2000 (first entry)

XX E. coli proliferation associated protein sequence SEQ ID NO:325.

DE Escherichia coli; E. coli; proliferation; inhibition; screening;

XX antimicrobial; bacterial growth; antisense therapy; antibacterial.

KW Escherichia coli.

OS WO200044906-A2.

XX 03-AUG-2000.

PF 27-JAN-2000; 2000WO-US02200.

XX 27-JAN-1999; 99US-0117405.

XX (ELIT-) ELITRA PHARM INC.

XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Frellich JM, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2000-514822/46.

DR N-PSDB; AAA65973.

XX Novel polynucleotides and polypeptides associated with microorganism

PT proliferation, used to identify inhibitors of bacterial growth and

11	189.5	7.9	148	22	AA15968	Putative P. abyssi
12	181.5	7.6	379	23	AAU91034	Herbicidally activ
13	178	7.4	306	22	AAU35442	Haemophilus influe
14	170	7.1	81	23	ABP33191	Human ORF2164 prot
15	169	7.1	309	22	AAU34813	E. coli cellular p
16	169	7.1	309	22	AAG98431	Escherichia coli p
17	167	7.0	310	22	AAU38383	Salmonella typhi c
18	165.5	6.9	315	23	ABP39676	Staphylococcus epi
19	165	6.9	309	22	AAU38290	Salmonella typhi c
20	165	6.9	401	21	AAG38354	Arabidopsis thalia
21	165	6.9	421	21	AAG09521	Arabidopsis thalia
22	162	6.8	309	18	AAW20221	H. pylori cytoplas
23	158	6.6	309	22	AAU34587	E. coli cellular p
24	156	6.5	729	22	ABG29904	Novel human diagno
25	154.5	6.5	302	22	AAG82984	S. epidermidis ope
26	154.5	6.5	352	21	AAG09068	Arabidopsis thalia
27	154	6.4	343	21	AAG38355	Arabidopsis thalia
28	154	6.4	363	21	AAG09522	Arabidopsis thalia
29	153	6.4	322	22	AAW38771	Human polypeptide
30	152	6.4	322	22	AAW38771	Human transfease
31	151.5	6.3	326	21	AAG09069	Arabidopsis thalia
32	150	6.3	137	23	ABP40715	Staphylococcus epi
33	148.5	6.2	306	22	ABP52865	Escherichia coli p
34	148.5	6.2	337	21	AAG38356	Arabidopsis thalia
35	148.5	6.2	357	21	AAG09523	Arabidopsis thalia
36	148	6.2	362	22	ABP61630	Drosophila melanog
37	147.5	6.2	303	23	ABP28656	Streptococcus poly
38	147	6.1	384	23	ABP91470	Herbicidally activ
39	146.5	6.1	306	22	ABP67588	Amino acid sequenc
40	146	6.1	206	22	ABP66042	Drosophila melanog
41	145	6.1	132	22	AAU33903	Staphylococcus aur
42	145	6.1	132	22	AAU36710	Staphylococcus aur
43	145	6.1	132	22	AAU37282	Staphylococcus aur
44	145	6.1	132	22	AAU37558	Staphylococcus aur
45	144.5	6.0	310	21	AAG09070	Arabidopsis thalia

proliferation, for use in antisense therapy -

Claim 11; Page 242-243; 316pp; English.

AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide sequences derived from *Escherichia coli* which inhibit *E. coli* proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent nucleotide and protein sequences associated with *E. coli* proliferation. AAA66056 and AAA66057 represent primers used for sequencing *E. coli* proliferation inhibiting nucleotide inserts in an example from the present invention. Methods from the present invention can be used to identify a proliferation- required gene in a microorganism, by contacting a microorganism with a proliferation- required gene activity inhibitory nucleic acid identified in another organism, and determining if inhibition occurs in the second microorganism. The nucleic acid sequences identified as being required for bacterial growth and proliferation, can be used for antisense therapy for killing bacteria.

Sequence 477 AA;

Query Match 100.0%; Score 2393; DB 21; Length 477;
Best Local Similarity 100.0%; Pred. No. 4.2e-208;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKVTLPFERAGVMVGDVMDLDRYWGTSRISPEAPVPVVKVNTIERPCGAANVAMNI 60
DB 1 MKVTLPFERAGVMVGDVMDLDRYWGTSRISPEAPVPVVKVNTIERPCGAANVAMNI 60
QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLVLSRNQQLRLD 120
DB 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLVLSRNQQLRLD 120
QY 121 FEEGFEVDVDPQLHERINQALSSIGALVSDYAKGALASVQOQIOLARKAGVPLIDPKG 180
DB 121 FEEGFEVDVDPQLHERINQALSSIGALVSDYAKGALASVQOQIOLARKAGVPLIDPKG 180
QY 181 TDFERYGATLLTPNLSEFEAVGCKTEEEIIVERGKMLIADYELSAALLVTRSEQGSLL 240
DB 181 TDFERYGATLLTPNLSEFEAVGCKTEEEIIVERGKMLIADYELSAALLVTRSEQGSLL 240
QY 241 QPKAPLHMPPTQAOEVYDVTGAGTIVGLAATLAAGNSLEEACFFANAAAGVVGKLG 300
DB 241 QPKAPLHMPPTQAOEVYDVTGAGTIVGLAATLAAGNSLEEACFFANAAAGVVGKLG 300
QY 301 STVSPIELENVAVRGRADTGFVMTTEELKLAVAAARKRGEKVMVTNGVFDILHAGHVSYL 360
DB 301 STVSPIELENVAVRGRADTGFVMTTEELKLAVAAARKRGEKVMVTNGVFDILHAGHVSYL 360
QY 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALEAVDVWVSFEEDTPQRL 420
DB 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALEAVDVWVSFEEDTPQRL 420
```

```
QY 421 IAGILPDLVKGGDYKPEETAGSKVWANGGEVLVLFNFDGCSSTNIKKIQDQKG 477
DB 421 IAGILPDLVKGGDYKPEETAGSKVWANGGEVLVLFNFDGCSSTNIKKIQDQKG 477
```

RESULT 2
AAU34708

AAU34708 standard; Protein; 477 AA.

AAU34708;

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #289.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Escherichia coli.

W0200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

PI WPI; 2001-611495/70.

XX N-PSDB; AAS52567.

DR New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

PT Example 3; Seq ID No 10301; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes themselves and the encoded proteins. The prokaryotes used are

CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 477 AA;

QY Query Match 100.0%; Score 2393; DB 22; Length 477;

DB Best Local Similarity 100.0%; Pred. No. 4.2e-208;

Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVTLPFERAGVMVGDVMDLDRYWGTSRISPEAPVPVVKVNTIERPCGAANVAMNI 60

DB 1 MKVTLPFERAGVMVGDVMDLDRYWGTSRISPEAPVPVVKVNTIERPCGAANVAMNI 60

QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLVLSRNQQLRLD 120

DB 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLVLSRNQQLRLD 120

QY 121 FEEGFEVDVDPQLHERINQALSSIGALVSDYAKGALASVQOQIOLARKAGVPLIDPKG 180

DB 121 FEEGFEVDVDPQLHERINQALSSIGALVSDYAKGALASVQOQIOLARKAGVPLIDPKG 180

QY 181 TDFERYGATLLTPNLSEFEAVGCKTEEEIIVERGKMLIADYELSAALLVTRSEQGSLL 240

DB 181 TDFERYGATLLTPNLSEFEAVGCKTEEEIIVERGKMLIADYELSAALLVTRSEQGSLL 240

QY 241 QPKAPLHMPPTQAOEVYDVTGAGTIVGLAATLAAGNSLEEACFFANAAAGVVGKLG 300

DB 241 QPKAPLHMPPTQAOEVYDVTGAGTIVGLAATLAAGNSLEEACFFANAAAGVVGKLG 300

QY 301 STVSPIELENVAVRGRADTGFVMTTEELKLAVAAARKRGEKVMVTNGVFDILHAGHVSYL 360

||||| 301 STVSPLEENAVGRADTGFVGMTEELKLAVAAARKRGEKVVMTNGVDFILHAGHSYL 360
Qy 361 ANARKLGRLIVAVNSDASTKRLGDSRPVNPLEORMIVLGALEAVDMVVSFEEDTPORL 420
Db 361 ANARKLGRLIVAVNSDASTKRLGDSRPVNPLEORMIVLGALEAVDMVVSFEEDTPORL 420
Qy 421 IAGILPDLLVKGDDYKPEIAGSKEVWANGGEVLVNFEDGCGSTTNIKKIQDQKKG 477
Db 421 IAGILPDLLVKGDDYKPEIAGSKEVWANGGEVLVNFEDGCGSTTNIKKIQDQKKG 477

RESULT 3
AAG98421
ID AAG98421 standard; Protein; 477 AA.
AC AAG98421;
XX
DT 21-SEP-2001 (first entry)
DE Escherichia coli protein sequence SEQ ID NO:469.
XX
KW Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition.
XX
OS Escherichia coli.
XX
PN W0200148209-A2.
XX
PD 05-JUL-2001.
XX
PF 19-DEC-2000; 2000WO-US34419.
XX
PR 23-DEC-1999; 99US-0173005.
XX
PA (ELIT-) ELITRA PHARM INC.
PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX
DR WPI; 2001-457376/49.
DR N-PSDB; AAH81477.
XX

Novel nucleic acids encoding proteins required for Escherichia coli proliferation, useful for screening for antimicrobial agents -
Example 4; Page 585-586; 596pp; English.

The present invention describes a purified or isolated nucleic acid sequence (I) consisting essentially of one of the 93 nucleotide sequences given in AAH81202 to AAH81294, where expression of the nucleic acid in a microorganism is capable of inhibiting proliferation of a microorganism. (I) have antibacterial and antibiotic activities, and can be used in gene therapy. Expression of (I) in a microorganism inhibits proliferation of the microorganism, and the manufactured antibiotic is useful for reducing the activity or level of a gene product required for proliferation of a microorganism in a subject, specifically humans. The nucleic acids that inhibit bacterial growth or proliferation can be used as antisense therapeutics for killing bacteria. In addition to therapeutic applications, the nucleic acid sequences complementary to sequences required for proliferation can be used as diagnostic tools. For example, nucleic acid probes complementary to proliferation-required sequences that are specific for particular species of microorganisms can be used as probes to identify particular microorganism species in clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent oligonucleotides, which are used in the exemplification of the present invention.

Sequence 477 AA;
Query Match 100.0%; Score 2393; DB 22; Length 477;
Best Local Similarity 100.0%; Pred. No. 4.2e-208;

Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKVTLPFERAGVGVVGDVMDLDRYWGPTSRISPPAPVPVVKVNTIEERPGGAANVAMNI 60
Db 1 MKVTLPFERAGVGVVGDVMDLDRYWGPTSRISPPAPVPVVKVNTIEERPGGAANVAMNI 60
Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTPTITKLRVLSRNOQLIRLD 120
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTPTITKLRVLSRNOQLIRLD 120
Qy 121 FEEGEGVDPQPLHERIQAALSSIGALVSDYAKALASVQOMIQARKAGVPVLIDPKG 180
Db 121 FEEGEGVDPQPLHERIQAALSSIGALVSDYAKALASVQOMIQARKAGVPVLIDPKG 180
Qy 181 TDFERYRGATLLTPNLSEFEAVVGKCTEEIEVERGMKLIADYELISALLVTRSEOGMSLL 240
Db 181 TDFERYRGATLLTPNLSEFEAVVGKCTEEIEVERGMKLIADYELISALLVTRSEOGMSLL 240
Qy 241 QPGKAPLHMTQAQEVYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVVVKLG 300
Db 241 QPGKAPLHMTQAQEVYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVVVKLG 300
Qy 301 STVSPLEENAVGRADTGFVGMTEELKLAVAAARKRGEKVVMTNGVDFILHAGHSYL 360
Db 301 STVSPLEENAVGRADTGFVGMTEELKLAVAAARKRGEKVVMTNGVDFILHAGHSYL 360
Qy 361 ANARKLGRLIVAVNSDASTKRLGDSRPVNPLEORMIVLGALEAVDMVVSFEEDTPORL 420
Db 361 ANARKLGRLIVAVNSDASTKRLGDSRPVNPLEORMIVLGALEAVDMVVSFEEDTPORL 420
Qy 421 IAGILPDLLVKGDDYKPEIAGSKEVWANGGEVLVNFEDGCGSTTNIKKIQDQKKG 477
Db 421 IAGILPDLLVKGDDYKPEIAGSKEVWANGGEVLVNFEDGCGSTTNIKKIQDQKKG 477

RESULT 4
AAU38166
ID AAU38166 standard; Protein; 477 AA.
XX
AC AAU38166;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #57.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS56025.
XX
PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 13759; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 477 AA;

Query Match 94.7%; Score 2265; DB 22; Length 477;
Best Local Similarity 93.5%; Pred. No. 1.7e-196;
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKVTLPEFERAGVMVGVDMVDYRWYGTPTSRISPEAPVPVVKVNTIEERPGGAANVAMNI 60
DB 1 MKVNLPAFERAGVMVGVDMVDYRWYGTPTCRISPEAPVPVVKVNTIEERPGGAANVAMNI 60
QY 61 ASLGANARLVGLTGIDDAARALSKSLADNVKCDVSVTPHTITKRLVLSRNOQLRLD 120
DB 61 ASLGANARLVGLTGIDDAARALSKSLAEVNVKCDVSVTPHTITKRLVLSRNOQLRLD 120
QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180
DB 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALTSVQTMISLARAGVPVLIDPKG 180
QY 181 TDFERYRGATLLTPNLSEFAVVGCKTEEEIVERGMKLIADYELLSALLVTRSEQGMSLL 240
DB 181 TDFERYRGATLLTPNLSEFAVVGCKTSEDELVERGMKLIADYDLSALLVTRSEQGMSLL 240
QY 241 QPKAPLHPTQAOEYVDYTGAGDTVIGVLAATLAAGNSLEERCFANAAAGVVGKLT 300
DB 241 QPNKAPLHPTQAOEYVDYTGAGDTVIGVLAATLAAGNTLEEACYPANAAAGVVGKLT 300
QY 301 STVSPTELENAVGRADTGFVMTTEELKLVAAAKRGEKVVMTNGVFDILHAGHVSYL 360
DB 301 STVSPTELENAVGRADTGFVMTTEELKLVAAAKRGEKVVMTNGVFDILHAGHVSYL 360
QY 361 ANARKLGDLRIIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420
DB 361 ANARKLGDLRIIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALESDVMVVSFEEDTPQRL 420
QY 421 IAGILPDLLVKGDDYKPEETAGSKVWANGGEVVLVNFEDGCGSTTNIKKIQODKK 476
DB 421 IAGILPDLLVKGDDYKPEETAGSEVWANGGEVVLVNFEDGCGSTTNIKKIQOTSE 476
RESULT 5
AAU36486
ID AAU36486 standard; Protein; 474 AA.
XX
AC AAU36486;
XX
DT 14-FEB-2002 (first entry)
XX Pseudomonas aeruginosa cellular proliferation protein #476.
DE
XX

KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54345.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12079; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 474 AA;
Query Match 56.4%; Score 1350.5; DB 22; Length 474;
Best Local Similarity 57.9%; Pred. No. 1.3e-113;
Matches 274; Conservative 70; Mismatches 128; Indels 1; Gaps 1;
QY 1 MKVTLPEFERAGVMVGVDMVDYRWYGTPTSRISPEAPVPVVKVNTIEERPGGAANVAMNI 60
DB 1 MKLSMRFDOAPVLYVGVDMVDYRWYGTPTSRISPEAPVPVVRVEQHEDRFGGAANVANI 60
QY 61 ASLGANARLVGLTGIDDAARALSKSLADNVKCDVSVTPHTITKRLVLSRNOQLRLD 120
DB 61 AALGAQALLVGVTRGDEADSLANSKAGVDARFORDISQPTIVKLVMSRHQQLLRVD 120
QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180
DB 121 FEEFPR-TDAAALAVDVESLAKVKVLSVSDYCKGALQNHQVLIQAARARNIPVLADPKG 179
QY 181 TDFERYRGATLLTPNLSEFAVVGCKTEEEIVERGMKLIADYELLSALLVTRSEQGMSLL 240
DB 180 KDPATYRGASLTIPNLSEFETIVGRCADAEALVAKGQALMSELDILGALLVTRGHEGMTLL 239

QY 241 QPKAPLHPTQAOEVYDVTGAGDTVIGVLAATLAAGNSLEAECEFFANAAAGVVVVKLGT 300
DB 240 RHGQPALHLPARAREVEDVTGAGDTVISTLAALAAEELPSAVGLANLAGIIVVVKLGT 299
QY 301 STVSPIELENAVRGADTGFQVMTTEEELKLAVAARKEGVVMTNGVDFDILHAGHVSYL 360
DB 300 AASAPELRAVQREGSERCIVGLEQLLAIEDARAHEKIVFTNGCFDILHAGHVTYL 359
QY 361 ANARKLGRLLIVANSASTKRLKDSRPVNPLEQRMIVGALCAVDMVVSFEEDTQRL 420
DB 360 EQARAQDRLLVGVNDASVTRLKGVGRPTNSVDRMAVLAGLAVDMVVSFAEDTPERL 419
QY 421 IAGTLPDLLVKGGBYKPEETAGSKEVWANGGEVLVLFNEDGCSCTNIIKKIQO 473
DB 420 LEQVRPDLVKGGBYGVQVGAQIVKAYGGEVRVLGVENSSTTAIVEKIRQ 472

RESULT 6
AAU72936
ID AAU72936 standard; Protein; 323 AA.
XX
AC AAU72936;
XX
DT 12-MAR-2002 (first entry)
XX
DE Neisseria meningitidis virulence protein #26.
XX
KW Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
KW infection; Gram-negative bacteria; antimicrobial.
XX
OS Neisseria meningitidis.
XX
PN WO200185772-A2.
XX
PD 15-NOV-2001.
XX
PF 08-MAY-2001; 2001WO-GB02003.
XX
PR 08-MAY-2000; 2000GB-0011108.
XX
PA (MICR-) MICROSCIENCE LTD.
XX
PI Tang C;
XX
DR WPI: 2002-066593/09.
DR N-PSDB: AAS97221.
XX
PT New peptide encoded by operon including virulence genes of Neisseria
PT meningitidis, useful as vaccine component for treating or preventing
PT meningitis and for identifying antimicrobial drug -
XX
PS Claim 4; Page 112-113; 423pp; English.
XX
CC The invention relates to a peptide (I) encoded by an operon (II) of
CC Neisseria meningitidis including virulence genes, or a related molecule
CC having a 40% sequence similarity at the peptide or nucleotide level in a
CC Gram-negative bacterium, or its functional fragment, for therapeutic or
CC diagnostic use. (I) and (II) are useful in the manufacture of a
CC medicament for treating or preventing a condition (e.g., meningitis)
CC associated with infection by Neisseria or Gram-negative bacteria. The
CC product is useful for veterinary treatment and in a screening assay for
CC the identification of an antimicrobial drug. The vaccines have
CC prophylactic applications. AAU72911-AAU73014 represent N. meningitidis
CC virulence proteins of the invention.
XX
SQ Sequence 323 AA;
Query Match 31.2%; Score 745.5; DB 23; Length 323;
Best Local Similarity 50.5%; Pred. No. 5.1e-59;
Matches 156; Conservative 49; Mismatches 103; Indels 1; Gaps 1;

QY 8 FERAGVMVGVMDLRYWYGPTSRISPEAPVPVAKIGRIDQACGAANVARNIASLGKV 67

DB 14 FAQAKVLVGVMDLRYWFGDVSRISPEAPVPVAKIGRIDQACGAANVARNIASLGKV 73
QY 68 RLVLGTGIDDAARALSCLADVNVKCDVSVPTHTTKRLVLSRNQOLIRLDFEEGEG 127
DB 74 GLLSVTGNDEAADALDALMVQDGVASVLMRDQKQIATTVKLRVARNQOLIRLDFEEHPNR 133
QY 128 VDPOLHERINOALSSIGALVLSYAKGALASVOQMIOQLARKAGVPVLIDPKGTDFFERYR 187
DB 134 EVLEQIKRKREIIPEDVDAIIFSDYGGGLSHISDMIDWAKHEGKTVLIDPKGDDYKYA 193
QY 188 GATLLTNLSFEAVVGVCKKTEEEIVERGMKLIADYELSALLVTRSEOGMSLLOPGKAPL 247
DB 194 GATLITPNRAELKEVGVSKNENDLTEKAQNLRRHLDTAILLTRSEEGMTLFSEGE-PI 252
QY 248 HMPQAOGEVDYTGAGDTVIGVLAATLAAGNSLEAECEFFANAAAGVVVVKLGTSTVSPIE 307
DB 253 YQPTRAQGEVDYSGAGDTVIAGMGLGLAAGCTMPMEAMYLANTAAGVVVAKLGTAVCSFAE 312
QY 308 LENAVERGRA 316
DB 313 LTKALSGQS 321

RESULT 7
AAW20763
ID AAW20763 standard; protein; 424 AA.
XX
AC AAW20763;
XX
DT 15-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 07apl1409orf4.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgaard BL;
XX
DR WPI: 1997-052306/05.
DR N-PSDB: AAT68016.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 61; Page 1174-1175; 1481pp; English.
XX
CC The present sequence is a Helicobacter pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

XX Sequence 424 AA;

Query Match 30.7%; Score 734.5; DB 18; Length 424;
Best Local Similarity 40.9%; Pred. NO. 7.6e-58;
Matches 177; Conservative 77; Mismatches 154; Indels 25; Gaps 9;

QY 51 GGAANYAMNIALSGANARLVGLTGIDDAARALSKSLADVNVKDFVSV-PTHPTTKLRV 109
DB 3 GGAANYANNTSLKAKVFLCGVGGDLKGGKHFISTLSIRIDTSGVLIDKTRCTTLKTRI 62

QY 110 LSRNQLIRLDEEGEGVDP-----OPLHERINQALSSIGALVLSDYAKGAL--ASVQ 161
DB 63 IAQNOQIVRVDEI---KDPLNADLRKNLDFIAEKIQEIDGVILSDYKNGVDFELTQ 118

QY 162 QMTQLARKAGVPLIDPKGTDFERYGATLLTPNLSEFAVVGKCKTEEEIVERGMKLTA 221
DB 119 TITLANKHKKLILCDPKCKDYKSHSLITPNRAELQALHLKLDSHANLSKALQILQ 178

QY 222 D-YELGALLVTRSEQMSLLQPKGAPLHPTQAEVYVDTGAGTVIGVLAATLAAGNSL 280
DB 179 EYTHIAMPLVTLSEQGIATLEKGEI-VNCPITIAKEVYDVTGAGTVIASLTLLESKSL 237

QY 281 EACFFANAAAGVVGKLGCTSVSPLENAVGRADTCFGVMTTEELKLAFAAAKRCGE 340
DB 238 KEACEFANAAAGVVGKMGSAAS---LEEIALINQHPKILPLEKL---LETLENNQ 291

QY 341 KYVMTNGVDFILHAGVSYLANARKLDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVL 400
DB 292 KIVFTNGCFDILKHGASHYLQAKALGDILVGLNSDNSIKRLKGDKRPVSEKDRFLL 351

QY 401 GALEAVDVVVSFEEDTPQRLIAGILPDILVKGDDYKPEIAGSKVWANGGEVLVNFED 460
DB 352 ASLSCVDYVWVFGEDTPIKLIQALKPDILVKGADYLNKEVIGSE----LAKETRLEFEE 407

QY 461 GCSTNNIKKIQ 473

DB 408 GYSTSAIEKIKR 420

RESULT 8

ID AAB07577 standard; Protein; 498 AA.

XX AAB07577;

AC AAB07577;

DT 20-OCT-2000 (first entry)

XX Protein encoded by the bleomycin (BLM) gene cluster ORF30.

DE BLM gene cluster; bleomycin gene cluster; polyketide metabolite;

XX bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;

KW thiazoline; bithiazoline; microbial metabolite; sugar.

XX Streptomyces verticillus.

XX Key Location/Qualifiers

FT Misc-difference 1

FT /note= "encoded by GTG"

XX WO2000040704-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00445.

PA (REGC) UNIV CALIFORNIA.

XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;

XX WPI; 2000-465974/40.

XX N-PSDB; AAA58471.

XX New bleomycin gene cluster components useful for peptide and/or

PS polyketide metabolites, especially bleomycin, production and for

XX chemically modifying biological molecules -

XX Disclosure; Page 97-137; 162pp; English.

XX AAB07556-78 represent proteins encoded by open reading frames (ORFs)
CC 8 to 30 of the BLM (bleomycin) gene cluster. The proteins encoded
CC by the gene cluster are useful for producing peptides and/or polyketide
CC metabolites, especially bleomycin or bleomycin analogues. They are
CC also useful for chemically modifying biological molecules to produce
CC branched methyl groups, and for coupling amino acids and fatty
CC acids. They may be reacted with an apo-carrier protein and coenzyme A
CC to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC domains can be used individually or collectively to produce
CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC microbial metabolites. The BLM gene cluster may also be used to produce
CC sugars.

XX Sequence 498 AA;

Query Match 25.5%; Score 610.5; DB 21; Length 498;

Best Local Similarity 37.7%; Pred. No. 1.7e-46;

Matches 182; Conservative 68; Mismatches 196; Indels 37; Gaps 15;

QY 13 VMVVGDMLDRYWGYGTSRISPEAPVPVVKVNTIEERPGGAAVNMNIALSGANARLVGL 72

DB 17 VLIVGVDILDTVWVGATSGLCRESPPAVTLTSVAHQCGGAANVANLGAEPVLSA 76

QY 73 TGIDDAARALSKSL--ADNVNKCDFVSVTPHTITIKRLVLSRNQQLIRLDFEGFEGVDP 130

DB 77 TGDGRAGRLREALRARDVDVTGGLFVQ-PCRTTIVTKRRVMACQMLRLD--EG--GEHP 131

QY 131 QP-----LHERINQALSSIGALVLSDYAKGALASVQMIOLA--RKAGVPVLI--DP 178

DB 132 LPVATDTGSRLLERAAAGLLPAVDVIVSDYGYG-VMEPDTVARLAARELGSTLVDSR 190

QY 179 KGTDFERYGATLLTPNLSEF-----EAVVGKCKTEEEIVERGMKLIADYELSALLVT 231

DB 191 RPARFTALR-ASAVKPNHAEALRLDAGEPPGPARA-DWAAALGDRLLRLTGAERVALT 248

QY 232 RSEQGMSLLQPGKAPLHPTQAE--VYDVTGAGDTVIGVLAATLAAGNSLEACFFANA 289

DB 249 LDADGSLLEFDRPPVPTFARGSRAPVTAAGVAGDAFTAALTTLAAGADSAVAABELASA 308

QY 290 AGVVGKLGSTVSPLEENAVGRADTCFGVMTTEE-ELKLAVAAARKGKGVWTVNGV 348

DB 309 AAGTAVATPGTSTWHADELRLGG---TGKVCRTGTLFARLDDPAARD--RWVFTNGC 363

QY 349 FDLHAGVSYLANARKLDRLIVAVNSDASTKRLKGDSPVNPVLEGRMIVLGAELAVDW 408

DB 364 FDLHGGHVSCLSRKELGDLVGVNSDSVRRKGPVPRPIAERMRVLAALSCVDL 423

QY 409 VVSFEEDTPORLIAGILPDILVKGDDYKPEETAGSKVWANGGEVLVNFEDGCSNTNII 468

DB 424 VVPFDDSDPAALTEALRPEVYAKGGDYTLATLPEALPVQRLGCVWHLPLSVADTSTTDII 483

QY 469 KKI 471

DB 484 RRI 486

RESULT 9

AAU36261

ID AUA36261 standard; Protein; 308 AA.

XX

AAU36261;
14-FEB-2002 (first entry)
Pseudomonas aeruginosa cellular proliferation protein #251.
Antisense: prokaryotic cellular proliferation protein;
antibiotic; antibacterial; drug design.
Pseudomonas aeruginosa.
WO200170955-A2.
27-SEP-2001.
21-MAR-2001; 2001WO-US09180.
21-MAR-2000; 2000US-191078P.
23-MAY-2000; 2000US-206848P.
26-MAY-2000; 2000US-207727P.
23-OCT-2000; 2000US-242578P.
27-NOV-2000; 2000US-253625P.
22-DEC-2000; 2000US-257931P.
16-FEB-2001; 2001US-269308P.
(ELIT-) ELITRA PHARM INC.
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
Yamamoto RT, Xu HH;
WPI; 2001-611495/70.
N-PSDB; AAS54120.
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Example 3; Seq ID No 11854; 51pp; English.
The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the
genes, their use in the discovery of novel antibiotics, the essential
genes themselves and the encoded proteins. The prokaryotes used are
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
invention is also useful for the identification of potential new targets
for antibiotic development. The antisense nucleic acids can also be used
to identify proteins used in proliferation, to express these proteins,
and to obtain antibodies capable of binding to the expressed proteins.
The proteins can be used to screen compounds in rational drug discovery
programmes. The antisense nucleic acid sequence is also useful to screen
for homologous nucleic acids which are required for cell proliferation in
a wide variety of organisms. The present sequence represents an
essential prokaryotic cellular proliferation protein.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.
Sequence 308 AA;
Query Match 8.2%; Score 197; DB 22; Length 308;
Best Local Similarity 26.5%; Pred. No. 2.5e-09;
Matches 86; Conservative 50; Mismatches 153; Indels 36; Gaps 11;
QY 10 RAGVNVGVGDVMDLRYWYGT-----SRISPEAPVPVVKVNTIERPGG-AANVAMNTASL 63
DB 2 QAKVLVVGSLNMDLVVRAPLPRGGETLAGQSFIV-----PGKGANQAAVAARL 52
QY 64 GANARLVGLTGIDDAARALSKSLADVNVKDFVS-VPTHPTITKRLVLSRNOQLRLDFE 122
DB 53 GAENVAMIGLCGDAYGQDLVRLAQEGIDCGVERVAGSSGVALLVDDSSQNAIVIVA 112
QY 123 EGFEGVDPQL--HERINQALSSITGALVLSDYAKGALASVQOMIQLARKAGVPLIDP-- 178

Db 113 GGNGHLSPAVLARHEHLEQ-----AQVVYQQLSEPLETVGHVLRRAHALCKTKVILNPAP 167
QY 179 --KGTDEYRGATLLTPNLSEFEAVVGKCKTEEEIVE---RGMKLTADYELSALLYTRS 233
Db 168 ATRDVPAAEWLPLVDYLVFNTESELL---CRLPVDLSLESAGRAAERUREMGAGRVIVTLG 224
QY 234 EQGMSLLQPGKAPLHMPTQAOEYVDVTGAGDTVIGVLAANTLAAGNSLEEACFFANAAAGV 293
Db 225 AQGALLVGEGRVE-HFPVARVKALDTTAAGDTFVGGAALARGLDDEAAAIIRFGQAAAI 283
QY 294 VVGKLTSTVSPI--ELENVAVGRA 316
Db 284 SVTRLGAQTSIPSRREVERALVGEA 308
RESULT 10
AAU35299
ID AAU35299 standard; Protein; 305 AA.
XX AC AAU35299;
XX DT 14-FEB-2002 (first entry)
XX DE Enterococcus faecalis cellular proliferation protein #586.
XX KW Antisense: prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Enterococcus faecalis.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS53158.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 10892; 51pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.


```

SQ Sequence 379 AA;
Query Match 7.6%; Score 181.5; DB 23; Length 379;
Best Local Similarity 25.6%; Pred. No. 8.6e-08;
Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

Qy 28 PTERISPEAPVVKVNT-----IEERP-----GGAANVAMNIASL 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 PKSAVDAHAP-PLVVGSANADIVYEIERLPKEGETISAKTGQTLAGGKGANQAACGAKL 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 64 GANARLVCLTGIDDAARALSKSLAD--VNVKDFV-SVPTHTP---ITKLRVLSRNOQLI 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119 MYPTFYVGRGLGDAHGKPIAEALGDGCGVHLDVRSYNNPTGHAVYMLSDGQNSI 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 118 RLDPEEGEGVD---POPLHERINOALSSIGALVLSDYAKGALASVOOQLARKAGVP 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 -----VGGANMKAWPEIMSDDDLEIVRNAGIVLQREIPDSINI--QVAKAVKAGVP 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 174 VLIDPKGDF---ERYRGATLLTPNLSEFPAVVGKCKTEEEIVERGMKLIADYELSALL 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 VILDVGGMDTPIPNELLDSDILSPNETELSLRGTGMPPTTETFEQISQAVAKCHKLGKQVL 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 230 VTRSEOGNSLLQPKAPLHMP-TQAEVYDVTGAGDVTIGVLAATLAAGNSLEERACFTAN 288
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 VKLGSKGSALFIQGEKPIQOSIIPAAQVYDTTGADTFTAAFAVAMVEGKSHEECLRFPA 349
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 289 AAAGVVVGKLGSTVSP 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 AAASLCVQVKAIPSP 366
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
AAU35442
ID AAU35442 standard; Protein; 306 AA.
XX
AC AAU35442;
XX
DT 14-FEB-2002 (first entry)
XX
DE Haemophilus influenzae cellular proliferation protein #83.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Haemophilus influenzae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS53301.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 11035; 511pp; English.
XX
XX
```

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acid can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 306 AA:

Query Match 7.4%; Score 178; DB 22; Length 306;
Best Local Similarity 22.7%; Pred. No. 1.3e-07;
Matches 66; Conservative 45; Mismatches 116; Indels 64; Gaps 7;

Qy 51 GGAANVAMNIASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLRLV 110
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 GKGANQAVAAARLGAFAVIFSGISDSIGTKMKNFAQEGID-----TTH-----IN 86
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 111 SRNQLIRLDFEGEGVDPOPLHERINOALSSIGALVLSDYAKGALASVOQMI----- 164
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 TVSQEMTGMF-----IQVAKSSENSIVL---ASGANSLSSEVVYRQSEA 128
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 165 -----QLARKAGVPVLIDPKGDF---ERYRGATLLTPNLSE 198
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 QIAQSDCLLMQLEPLSGVELAAQIAKKNVKNVLPAPQAQLSLDSLIDITPNETE 188
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 199 FEAVVGKCKTEEEIVERGMKLIADYELSALLVTRSEOGNSLLQPKAPLHMPQAEVYD 258
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 AEILTGVEVADEQSAVKAASFVHDKGITVMTLGAKGVFVSRKGRKRIKGCVCQAI-D 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 259 VTGAGDVTIGVLAATLAAGNSLEEAFCFAANAAGVVVGKLGSTVTSPIELE 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 248 TTAAGDTFNGGFVTALLEEKSFDEAIRFGQAAAAISVTKKGAQSSIPTKQE 298
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
ABP33191
ID ABP33191 standard; Protein; 81 AA.
XX
AC ABP33191;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF2164 protein, SEQ ID NO:4328.
XX
KW Human; ORF; open reading frame; ORF; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
XX
OS Homo sapiens.
XX


```
Db 58 IAFIACGTGDDSIGESVRQQLATDNI-----DITPVSVIKGESTGVALIFVNG-E 105
Qy 127 GVDPOPLHERINQALS-----SIGALVLSYAKGALASVQOMIQIOLARKAGVPV 174
Db 106 GENVIGIHAGANAALSPALVEAQRERIANASALLMQ--LESPLESVMAAAKIAHQNKTI 163
Qy 175 LIDP---KGTDPERYRGATLLTPNLSEFEAVVVGKCKTEEEIVERGMKLIADYELSALVT 231
Db 164 ALNPAPARELPDELLALVDIIITPNETAEKLTGIRVENDEDAKAAQVLHEKGIRTVLIT 223
Qy 232 RSEQMSLLQPKAPLHMPAQAEVYDVTGAGDVTIGVLAATLAAGNSLEEACFFANAAA 291
Db 224 LGSRGVWASVNGEGO-RVPGFRVQAVDTIAAGDTFNGALITALLEEKPLPEAIRFAHAA 282
Qy 292 GWVVGKLGTSVSP 305
Db 283 AIAVTRKGAQPSVP 296
```

Search completed: November 24, 2002, 23:15:49
Job time : 84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2002, 23:13:04 ; Search time 51 Seconds
(without alignments)
899.140 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERAGVMVGVDM.....FEDGCSFTNIIKKIQDKKG 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2393	100.0	477	1 B65093	ADP-heptose syntha
2	2385	99.7	477	2 G91120	ADP-heptose syntha
3	2385	99.7	477	2 F85965	probable kinase 24
4	2265	94.7	477	2 AG0891	ADP-heptose syntha
5	2076	86.8	476	2 A10080	ADP-heptose syntha
6	1712	71.5	476	1 C64127	ADP-heptose syntha
7	1350.5	56.4	474	2 A83022	LPS biosynthesis p
8	904.5	37.8	483	2 F87700	rfaE protein [impo
9	863	36.1	461	2 H81319	probable ADP-hept
10	819.5	34.2	463	2 C71887	probable adp-d-gly
11	811	33.9	315	2 G84936	kinase [imported]
12	809.5	33.8	461	1 B64627	ADP-heptose syntha
13	753.5	31.5	323	2 H81154	ADP-heptose syntha
14	745.5	31.2	323	2 H81951	probable DP-heptos
15	594	24.8	463	2 T34841	probable bifunctio
16	584.5	23.6	315	2 H70350	ADP-heptose syntha
17	391.5	16.4	157	2 H70317	glycerol-3-phospha
18	348.5	14.6	168	2 C81009	aut protein NMB207
19	340.5	14.2	168	2 G82031	probable transfera
20	311.5	13.0	164	2 I39548	aut protein - Alca
21	201.5	8.4	319	2 A83882	fructokinase BH185
22	197	8.2	308	2 B83403	ribokinase PAL950
23	194	8.1	299	2 E72311	ribokinase - Therm
24	192.5	8.0	298	2 AH0947	probable sugar kin
25	189.5	7.9	148	2 A75041	probable glycerol-
26	188	7.9	307	2 AB2029	ribokinase [import
27	184.5	7.7	293	1 D69690	ribokinase (EC 2.7
28	184.5	7.7	303	2 B87367	ribokinase [import
29	183.5	7.7	319	2 AC0942	probable carbohydr

30	181.5	7.6	148	2 H71120	hypothetical prote
31	181.5	7.6	378	2 F86307	hypothetical prote
32	179.5	7.5	383	2 D75288	carbohydrate kinas
33	179	7.5	137	2 A69427	glycerol-3-phospha
34	178	7.4	306	2 B64073	ribokinase (EC 2.7
35	176	7.4	294	2 H84115	ribokinase rbsk [1
36	175	7.3	311	2 E75385	carbohydrate kinas
37	174	7.3	303	2 F84401	hypothetical prote
38	174	7.3	316	2 A96952	fructokinase [impo
39	173.5	7.3	300	2 E86076	probable kinase y1
40	173.5	7.3	300	2 F91229	probable kinase [1
41	170.5	7.1	305	2 T44955	ribokinase (EC 2.7
42	170.5	7.1	343	2 D75260	probable carbohydr
43	170	7.1	300	2 A75599	ribokinase - Deino
44	169	7.1	309	1 KIECRB	ribokinase (EC 2.7
45	169	7.1	309	2 F91215	ribokinase [import

ALIGNMENTS

RESULT 1

B65093

ADP-heptose synthase homolog - Escherichia coli (strain K-12)

N:Alternate names: hypothetical protein b3052

C:Species: Escherichia coli

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002

C:Accession: B65093

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B65093

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-477 <BLAT>

A:Cross-references: GB:AB000387; GB:U00096; NID:gl789431; PIDN:AAC76088.1; PID:gl7894

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: hypothetical protein b3052

Query Match 100.0%; Score 2393; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 1.1e-144;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTLPEFERAGVMVGVDMVLDVWYGTSTRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60

Db 1 MKVTLPEFERAGVMVGVDMVLDVWYGTSTRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTPTITKLVLSRNOQLRLD 120

Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTPTITKLVLSRNOQLRLD 120

Qy 121 FEFGFEGVDPPLHERINQALSSIGALVSLDYAKGALASVQOMIQARAGVVPVLDPKG 180

Db 121 FEFGFEGVDPPLHERINQALSSIGALVSLDYAKGALASVQOMIQARAGVVPVLDPKG 180

Qy 181 TDFERYRGATLLTPNLSEFEAVVGKCTEEETIVERGKMLIADYELSAALLVTRSEQMSLL 240

Db 181 TDFERYRGATLLTPNLSEFEAVVGKCTEEETIVERGKMLIADYELSAALLVTRSEQMSLL 240

Qy 241 QPKAPLHMPYQAOQVYDVDTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300

Db 241 QPKAPLHMPYQAOQVYDVDTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300

Qy 301 STVSPIELENVVRGADTGFVGMTTEELKLVAAVAAARKKEKVVMTNGVFDILHAGHSVL 360

Db 301 STVSPIELENVVRGADTGFVGMTTEELKLVAAVAAARKKEKVVMTNGVFDILHAGHSVL 360

Qy 361 ANARKLGRLIVAVNSDASTKRLKGDSPVNPVLEORMIVLGALEAVDVVVSFEEDTPQRL 420

Db 361 ANARKLGRLIVAVNSDASTKRLKGDSPVNPVLEORMIVLGALEAVDVVVSFEEDTPQRL 420

Qy 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477
|||||
Db 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477

RESULT 2

G91120
ADP-heptose synthase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G91120
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen-
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G91120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB37358.1; PID:g13363408; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS3935
C:Superfamily: hypothetical protein b3052

Query Match 99.7%; Score 2385; DB 2: Length 477;
Best Local Similarity 99.6%; Pred. No. 3.6e-144;
Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKVTLPFERAGVMVGVDMVLDYWGTPSRISPEAPVPVVKVNTIEERPGGAANVANI 60
|||||
Db 1 MKVTLPFERAGVMVGVDMVLDYWGTPSRISPEAPVPVVKVNTIEERPGGAANVANI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTITIKLRVLSRNOQLRLD 120
|||||
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTITIKLRVLSRNOQLRLD 120

Qy 121 FEEGEGVDPPQPLHERINOALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
|||||
Db 121 FEEGEGVDPPQPLHERINOALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180

Qy 181 TDFERYGATLTPNLSEFAVVGKCTEEIEVERGMKLIADYELSAVVTRSEQMSLL 240
|||||
Db 181 TDFERYGATLTPNLSEFAVVGKCTEEIEVERGMKLIADYELSAVVTRSEQMSLL 240

Qy 241 QPKAPLHPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEACFFANAAAGVVGKLT 300
|||||
Db 241 QPKAPLHPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEACFFANAAAGVVGKLT 300

Qy 301 STVSPTELENVAVGRADTGFVMTTEELKLAFAAARKKEKVVMTNGVFDILHAGHVSYL 360
|||||
Db 301 STVSPTELENVAVGRADTGFVMTTEELKLAFAAARKKEKVVMTNGVFDILHAGHVSYL 360

Qy 361 ANARKIGDRLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALAVDWWVVSFEEDTPQRL 420
|||||
Db 361 ANARKIGDRLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALAVDWWVVSFEEDTPQRL 420

Qy 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477
|||||
Db 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477

RESULT 3
F85965
probable kinase Z4405 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85965
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <STO>
A:Cross-references: GB:AF005174; NID:g12517628; PIDN:AAG58186.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z4405
C:Superfamily: hypothetical protein b3052

Query Match 99.7%; Score 2385; DB 2: Length 477;
Best Local Similarity 99.6%; Pred. No. 3.6e-144;
Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKVTLPFERAGVMVGVDMVLDYWGTPSRISPEAPVPVVKVNTIEERPGGAANVANI 60
|||||
Db 1 MKVTLPFERAGVMVGVDMVLDYWGTPSRISPEAPVPVVKVNTIEERPGGAANVANI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTITIKLRVLSRNOQLRLD 120
|||||
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTITIKLRVLSRNOQLRLD 120

Qy 121 FEEGEGVDPPQPLHERINOALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
|||||
Db 121 FEEGEGVDPPQPLHERINOALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180

Qy 181 TDFERYGATLTPNLSEFAVVGKCTEEIEVERGMKLIADYELSAVVTRSEQMSLL 240
|||||
Db 181 TDFERYGATLTPNLSEFAVVGKCTEEIEVERGMKLIADYELSAVVTRSEQMSLL 240

Qy 241 QPKAPLHPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEACFFANAAAGVVGKLT 300
|||||
Db 241 QPKAPLHPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEACFFANAAAGVVGKLT 300

Qy 301 STVSPTELENVAVGRADTGFVMTTEELKLAFAAARKKEKVVMTNGVFDILHAGHVSYL 360
|||||
Db 301 STVSPTELENVAVGRADTGFVMTTEELKLAFAAARKKEKVVMTNGVFDILHAGHVSYL 360

Qy 361 ANARKIGDRLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALAVDWWVVSFEEDTPQRL 420
|||||
Db 361 ANARKIGDRLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALAVDWWVVSFEEDTPQRL 420

Qy 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477
|||||
Db 421 IAGILPDLVKGVDYKPEIAGSKVWANGGEVLVLPNFDGCGSTTNIKKIQODKKG 477

RESULT 4

AG0891
ADP-heptose synthase (EC 2.7.-.-) [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0891
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AG0891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07725.1; PID:g16504277; GSPDB:GN00176

Query Match 94.7%; Score 2265; DB 2: Length 477;

QY 427 DLLVKGDDYKPEETAGSKEWANGGEVLVNFEDGCGSTTNIKKIQ 472
|||||
Db 427 DLLVKGDDYKPEETAGSKEWANGGDYKVLNFENGCGSTTNIKKIK 472

RESULT 7

A83022
LPS biosynthesis protein RfaE PA4996 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83022

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A:Reference number: A82950; PMID:20437337; PMID:10984043

A:Accession: A83022

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-474 <STO>

A:Cross-references: GB:AE0040912; GB:AE004091; NID:99951274; PIDN:AAG08381.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: rfaE; PA4996

C:Superfamily: hypothetical protein b3052

Query Match 56.4%; Score 1350.5; DB 2; Length 474;

Best Local Similarity 57.9%; Pred. No. 2.1e-78;

Matches 274; Conservative 70; Mismatches 128; Indels 1; Gaps 1;

QY 1 MKYTLPEFRAGVMYGVDMLDYRWYGTSPISPEAPVPVVKVNTTEERPGGAANVAMNI 60

|||||
Db 1 MKLSMPRFDAQVLVYGVDMYDWRWYGTSPISPEAPVPVRYEQHEDRPGGAANVAMNI 60

QY 61 ASIGANARLVGLTGIDDAARALSKSLADNVNKCDFVSVPHTPTITKLRLVSRNQILRLD 120

|||||
Db 61 AALGAQALLVGVTRDEAADSLSLKAAGVDARFORIDSQPTIVKLRLVMSRHQQLLRVD 120

QY 121 FESEFEGVDPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQIOLARKAGVPVLDPKG 180

|||||
Db 121 FESEFPR-TDAAALADVESLAKVVLVDYKGALONHQVLQIQAARNIPVLADPKG 179

QY 181 TDFERYRGATLPLNLSPEFVAVYCKTEBEEIYVERGMKLIADYELSALLVTRSEOGMSLL 240

|||||
Db 180 KDEAIYRGASLITPLNLSPEFTIYVGRCADEAEELVAKGQALMSLDELGALLVTRGEHGTLL 239

QY 241 QPKGAPLHMPYTAQOEYDVTGAGDTVIGVLAATLAAGNSLEAEACFTANAAAGVVGKLG 300

|||||
Db 240 RHGQPALHLPARAREVPDVTGAGDTVISTLAAALAAAGEELPSAVGLANLAAGIVVGKLG 299

QY 301 STVSPTELENAVGRADTGFVMTTEELKLAFAAARKGKVVMTNGVFDILHAGHVSYL 360

|||||
Db 300 AATSAPELRAVOREGSEKRVGLGLQOLLAIEDARAHGEKIYFTNGCFDILHAGHVTYL 359

QY 361 ANARKGLDRLIVAVNSDASTKRLKGRSPVNPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420

|||||
Db 360 EQARAQGDRLIVGNDASVTRKLGVRPINSVDRMAVLAGLAVDWMVVSFEADTPERL 419

QY 421 IAGILPDLVKGDDYKPEETAGSKEWANGGEVLVNFEDGCGSTTNIKKIQ 473

|||||
Db 420 LEQVRPDLVKGDDYGVGEVQVGAQIKVAYGGEVRLVGLVENSSTTAIVEKIRQ 472

RESULT 8

F87700

rfaE protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: F87700

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; PMID:21173698; PMID:11259647

A:Accession: F87700

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-483 <STO>

A:Cross-references: GB:AE005673; NID:g13425394; PIDN:AAK25602.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3640

C:Superfamily: hypothetical protein b3052

Query Match 37.8%; Score 904.5; DB 2; Length 483;

Best Local Similarity 45.1%; Pred. No. 5e-50;

Matches 214; Conservative 66; Mismatches 186; Indels 9; Gaps 8;

QY 5 LPE-FERAGVMYGVDMLDYRWYGTSPISPEAPVPVVKVNTTEERPGGAANVAMNIASL 63

|||||
Db 8 LPRAFGKTVLVLDGVMDLDRFIYCAVDRIISPEAPVPVIAVEKETAMLGAGAGNARNVAAL 67

QY 64 GANARLVGLTGIDDAARALSKSL-ADNVNKCDFVSVPHTPTITKLRLVSRNQILRLDFE 122

|||||
Db 68 GAKAVLIGLVGRDDAGAALRGMDIAEAGLEAEALWDPARETTEKRVYISGHOMLRVDRE 127

QY 123 EGFEQVDPQPLHERINQALSSIGALVLSDYAKGAL--ASVQOMIQIOLARKAGVPVLDPKG 180

|||||
Db 128 DRSPG-DGAALLAFAFETRLASADVVLSDYAKGVLPVAVVRGALDAKAKAGKPVIVDPKS 186

QY 181 TDFERYRGATLPLNLSPEFVAVYCKTEBEEIYVE-RGMKLIADYE-LSALLVTRSEOGMS 238

|||||
Db 187 RDPARYDGATLIKPRNKEAAEATGIVETSDAASEDAGAAILAMAPGLQAALITRGAGMT 246

QY 239 LLOPGKAPLHMPYTAQOEYDVTGAGDTVIGVLAATLAAGNSLEAEACFTANAAAGVVGKL 298

|||||
Db 247 LAVRNPPHLPATAIEVFDVSGAGDTVAATLALVAAGASLAAQAOLANLAGLVVAKL 306

QY 299 GTSTVSPTELENAVGRADTGFVMTTEELKLAFAAARKGKVVMTNGVFDILHAGH 356

|||||
Db 307 GTDVTVAEELTACASSAQGEPGEIKIADREQAQRIEVEGWRARGLVKVGFTNGCFDILLHPGH 366

QY 357 VSLANARKGLDRLIVAVNSDASTKRLKGRSPVNPLEQRMIVLGALEAVDWMVVSFEEDT 416

|||||
Db 367 VSLLSQAQACADRLIVGLNLTDAVSKLGPTRVQEQGRATVLASLSSVDLVVLPDEDT 426

QY 417 PQRLIAGILPDLVKGDDYKPEETAGSKEWANGGEVLVNFEDGCGSTTNIKKI 471

|||||
Db 427 PLELIKAFRDVLVKGADYIVETVVGSDVVLVYGGKVVLAELKQOGSTTNLIARM 481

RESULT 9

H81319

probable ADP-heptose synthase Cj1150c [imported] - Campylobacter jejuni (strain NCTC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: H81319

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chl

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; PMID:20150912; PMID:10688204

A:Accession: H81319

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73404.1; PID:g696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

C:Superfamily: hypothetical protein b3052

Query Match 36.1%; Score 863; DB 2; Length 461;

Best Local Similarity 43.1%; Pred. No. 2e-47;

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000

A:Accession: B64627
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.B.; Kelley, J.M.; Cotton, M.D.; Meldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.; A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: B64627
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-461 <TOM>
A:Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07904.1; PID:g231399
C:Superfamily: hypothetical protein b3052

Query Match 33.8%; Score 809.5; DB 1; Length 461;
Best Local Similarity 40.6%; Pred. No. 5.1e-44;
Matches 191; Conservative 88; Mismatches 165; Indels 27; Gaps 9;

QY 13 VNVGVDMVDYWGPTSRISPEAPVPVVKVNTIERPFGAANVAMNIASLGANRLVGL 72

DB 4 ILVGLDADYLVWCKSERLSPAPVPVLEQRESKNLGGAAVANNLSLKAKVFLGV 63

QY 73 TGIDDAARALSGLADVNVKDFVSV-PTHPTITKLRLVSRNQQLRLDFEGFEGVDP- 130

DB 64 VGDLEGEHFISALKARGIDASGILIDKTRCTTLKTRIAQNOQIARVDKEI- ---KDPL 119

QY 131 -----OPLHERINQALSSIGALVSDYAKGAL-ASVQMIQIARKAGVPVLIDPKGTDF 183

DB 120 NADLRKLLDFTTEKIQEIDGVILSDYNGVLDLDFELTQAMIALANQHHLKLLCDPKGDKY 179

QY 184 ERYRGATLTPNLSEFEAVVGCKTETEEIVERGMKLIAD-YELSALLVTRSEQMSLLQP 242

DB 180 SKYSHASLTPTNTELEHALHLKLDHANLSKALQILKETYHIAMPLVTLSEQIAFLK 239

QY 243 GKAPLHMPQAEVVDVTGAGTIVGLVLAATLAAGNSLEECFFANAAAGVVGKLGST 302

DB 240 GEL-VNCPITAEVVDVTGAGTIVASLTLSLESMLKDADEFANAAAVVVGKMSAL 298

QY 303 VSPLENAVGRADTFCVMTTEELKLAFAAARKRGEKVVNTNGVFDILHAGHSYLAN 362

DB 299 AS---LEETALILNQTHPKILSLEKLETLTLDQ-----QXIIITNGCFDILLKHGHSYLOK 350

QY 363 ARKLGRILVAVNSDASTKRLKGRSPVNPFLQRMIVLGALEAVDMVWSFEEDTPORLTA 422

DB 351 AKALGDLILVGLNSDASIKRLKGRPIVSEKDRAPFLASLSCVDVYVVFEDTPIKLIQ 410

QY 423 GILPDLVLKGGDYKPEIAGSKSEVWANGGEVLVLPFEDGCSNTNIIKKIQ 473

DB 411 ALKPDILVRGADYLNKEVIGSE----FAKETHLMFEFGYSTAIEIKIR 457

RESULT 13

H81154

ADP-heptose synthase, probable NMB0825 [imported] - *Neisseria meningitidis* (strain MC58)
C:Species: *Neisseria meningitidis*

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: H81154

R:Rettlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: H81154

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <DET>

A:Cross-references: GB:AE002435; GB:AE002098; NID:g7226049; PIDN:AAF41238.1; PID:g722606

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB0825

Query Match 31.5%; Score 753.5; DB 2; Length 323;
Best Local Similarity 51.5%; Pred. No. 1.1e-40;
Matches 159; Conservative 49; Mismatches 100; Indels 1; Gaps 1;

QY 8 FERAGVWVGVDMVDYWGPTSRISPEAPVPVVKVNTIERPFGAANVAMNIASLGANA 67

DB 14 FAQAKVLVGVDMVDYWGPTSRISPEAPVPVAKIGRIDORAGGAANVARNIASLGGA 73

QY 68 RLVLGTGIDDAARALSGLADVNVKDFVSVPTHPTITKLRLVSRNQQLRLDFEGFEG 127

DB 74 GLLSVTGNDEADALDALMVQDGVASYLMRKQIATTVTKLVRVARNQQLRLDFEEHPNC 133

QY 128 VDPQPLHERINQALSSIGALVSDYAKGALASVQMIQIARKAGVPVLIDPKGTDFERYR 187

DB 134 EVLEQIKQYREILPEYDALIIFSDYKGGLSHISDMIDWAKHAGKTVLIDPKGDDYERYV 193

QY 198 GATLITPNLSEFEAVVGCKTETEEIVERGMKLIADYELSALLVTRSEQMSLLQPKAPL 247

DB 194 GATLITPNRAELKEVVGSKNESELTERKAQNLRRLDUTAVLLTTRSEQMTLSEGE-PI 252

QY 248 HMPTQAEVVDVTGAGTIVGLVLAATLAAGNSLEECFFANAAAGVVGKLGSTVSPIE 307

DB 253 YQPTRAQEVVDYSGADIVIAGMGLGLAAGCTMPEAMVLANTAAGVAVKLGCTAVCSFAE 312

QY 308 LENAVRGRA 316

DB 313 LIKALSGQS 321

RESULT 14

H81951

probable DP-heptose synthetase (EC 2.7.-.-) NMA1034 [imported] - *Neisseria meningitidis*
C:Species: *Neisseria meningitidis*
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: H81951

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: H81951

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <PAR>

A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84303.1; PID:g737

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: rfaE, NMA1034

C:Keywords: phosphotransferase

Query Match

Best Local Similarity 31.2%; Score 745.5; DB 2; Length 323;

Matches 156; Conservative 49; Mismatches 103; Indels 1; Gaps 1;

QY 8 FERAGVWVGVDMVDYWGPTSRISPEAPVPVVKVNTIERPFGAANVAMNIASLGANA 67

DB 14 FAQAKVLVGVDMVDYWGPTSRISPEAPVPVAKIGRIDORAGGAANVARNIASLGKV 73

QY 68 RLVLGTGIDDAARALSGLADVNVKDFVSVPTHPTITKLRLVSRNQQLRLDFEGFEG 127

DB 74 GLLSVTGNDEADALDALMVQDGVASYLMRKQIATTVTKLVRVARNQQLRLDFEEHPNR 133

QY 128 VDPQPLHERINQALSSIGALVSDYAKGALASVQMIQIARKAGVPVLIDPKGTDFERYR 187

DB 134 EVLEQIKQYREILPEYDALIIFSDYKGGLSHISDMIDWAKHAGKTVLIDPKGDDYERYV 193

QY 188 GATLITPNLSEFEAVVGCKTETEEIVERGMKLIADYELSALLVTRSEQMSLLQPKAPL 247

DB 194 GATLITPNRAELKEVVGSKNESELTERKAQNLRRLDUTAVLLTTRSEQMTLSEGE-PI 252

QY 248 HMPTQAEVVDVTGAGTIVGLVLAATLAAGNSLEECFFANAAAGVVGKLGSTVSPIE 307

Db 253 YQPTRAQEVYDVSAGDVTIAGMGLAAGCTMPEAMYLANTAAGVYVAKLGTAVCSFAE 312
Qy 308 LENA VRGA 316
Db 313 LTKALSGQS 321

RESULT 15

T34841

Probable bifunctional synthase /transferase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000

C:Accession: T34841

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999

A:Reference number: Z21559

A:Accession: T34841

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-463 <OLI>

A:Cross-references: EMBL:AL035478; PIDN:CAB36595.1; GSPDB:GN00070; SCOEDB:SC2G5.08

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC2G5.08

C:Superfamily: hypothetical protein b3052

Query Match 24.8%; Score 594; DB 2; Length 463;
Best Local Similarity 36.3%; Pred. No. 2.6e-30;
Matches 173; Conservative 71; Mismatches 189; Indels 44; Gaps 13;

Qy 9 ERAGVMVYGVYDVRISPEAPVPVVKVNTTEERPGGAAVAMNIASLGANAR 68
Db 10 DRTPLVVGGDALLDLDTGTADRIAPDPVPVQECARIRPGGAALAAAYLAARDGREVT 69
Qy 69 LVGLTGIDDAARALSKSLADYVNVKDFVSVTPHTI-TKLRLVLSRNOQLIRLDFEEGEG 127
Db 70 LIAGVGEDPAGLALRELLAP---WLKLIPLTGTVPKTRVLAQDRPVVRLDRGGG--- 123
Qy 128 VDPQLHERINQALSSIG---ALVLSDYAKGALASVQOMIOLARKAGVPLIDPKGTDFE 184
Db 124 ----RVREATDEARDALCARAVLVSDYGRGAADALRDVL----AARPLVMDPHPRGGP 175
Qy 185 RYRGATLLTPNLSEFAEVVVGKCKTE-----EEIVERGMKLIADYELSLVTRSEOGM 237
Db 176 PVPGTRLVTP--AEKEA-HGFAPSEGRPGGGLRAAALAAALVRDVRVAAVTVTLGSRG- 231
Qy 238 SLLQPGKAPLHMQAQEVYDVTGAGDVTIGVLAATLAAGNSLSEACFFANAAAGVVVGK 297
Db 232 ALLSYGEHPLLVAPAHAHGGSCGAGDRFAATAAGLLADGALVGEAVEGAYGAATAFYAA 291
Qy 298 LGTSTVSPIELENAVRGRADTGFVMTTEELKLAFAAAARKGE--KVYMTNGVFDILHAG 355
Db 292 GGAAPVPAGSERALALPDTD-----DPGALAARIAREHGTVVAAAGCFDILLHAG 342
Qy 356 HVSYLANARKLGDRILIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALEAVDWVVSFEED 415
Db 343 HVGLLQAAARRLGDCLVVCVNSDASVRRGKGGGRPVNPLADRVRLALACVDAAVDFED 402
Qy 416 TPORLIAGILPDLLVKGDDYKPEET--AGSKEVWANGGEVLVNFEDCCSTTNIKK 470
Db 403 TPERLLGELRPDVMVYGGDYAGADLPEAGLLKEW--GQAVLLPYLDGRSRSTALLAR 457

Search completed: November 24, 2002, 23:19:33

Job time : 53 secs


```

QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVTPHTTITKRLVLSRNLRLD 120
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVTPHTTITKRLVLSRNLRLD 120
QY 121 FEGFGVDPQPUHERINQALSSIGALVLSYAKGALASVQOQMIQARKAGVPLIDPKG 180
Db 121 FEGFGVDPQPUHERINQALSSIGALVLSYAKGALASVQOQMIQARKAGVPLIDPKG 180
QY 181 TDFERYRGATLLTPNLSEFAVGVCKTKEEIVERGMKLIADYELSALLVTRSEQMSLL 240
Db 181 TDFERYRGATLLTPNLSEFAVGVCKTKEEIVERGMKLIADYELSALLVTRSEQMSLL 240
QY 241 QPKAPLHPTQAEYVDVTGAGDTVIGVLAATLAAGNSLEACFPFANAAGVGVCKLGT 300
Db 241 QPKAPLHPTQAEYVDVTGAGDTVIGVLAATLAAGNSLEACFPFANAAGVGVCKLGT 300
QY 301 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKGKVVMTNGVDFILHAGHVSYL 360
Db 301 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKGKVVMTNGVDFILHAGHVSYL 360
QY 361 ANARKLGDRLLIVAVNSDASTKRLKGDSPVNPLEQRMVILGALEADVWVSEEDTPQRL 420
Db 361 ANARKLGDRLLIVAVNSDASTKRLKGDSPVNPLEQRMVILGALEADVWVSEEDTPQRL 420
QY 421 IAGILPDLLVKGDDYKPEELAGSKVWANGGEVLVLFNFDGCGSTTNIKKIOODKKG 477
Db 421 IAGILPDLLVKGDDYKPEELAGSKVWANGGEVLVLFNFDGCGSTTNIKKIOODKKG 477

RESULT 2
ID RFAE_HAEIN STANDARD; PRT; 476 AA.
AC O05074;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADP-heptose synthase (EC 2.7.-.-).
GN RFAE OR WAAE OR H1526.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32828; AAC23172.1; -
CC HSSP: P27623; IC02.
CC TIGR: H11526; -.
DR InterPro: IPR004821; Cytidylyltransf.

```

```

DR InterPro: IPR004820; Cytidylyltransf.
DR InterPro: IPR002173; pfkb.
DR Pfam: PF00294; pfkb; 1
DR Pfam: PF01467; Cytidylyltransf; 1.
DR TIGR: TIGR00125; cyt_tran_rel; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; FALSE_NEG.
DR PROSITE: PS00584; PFKB_KINASES_2; FALSE_NEG.
KW Lipopolysaccharide biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 476 AA; 51945 MW; 4F241C08D2C6951E CRC64;

Query Match 71.5%; Score 1712; DB 1; Length 476;
Best Local Similarity 71.0%; Pred. No. 5.2e-95;
Matches 331; Conservative 58; Mismatches 77; Indels 0; Gaps 0;

QY 7 EFERAGVWVGVMDLRYWYGTSPTRISPEAPVVPVVKVNTIEERPGGAANVAMNIALGAN 66
Db 7 EFKQAKVLVGLDVMYDFWFGATNRISPEAPVVPVVKVQNEERAGGAANVAMNIALNVP 66
QY 67 ARLVGLTGIDDAARALSKSLADVNVKCDVSVTPHTTITKRLVLSRNLRLDPEEGFE 126
Db 67 VQLMGLIGQDETSALSLLEKQIDCNFVALETHPTITKRLLSRHOQLLRDFEDFN 126
QY 127 GVDPOPLHERINOALSSIGALVLSYAKGALASVQOQMIQARKAGVPLIDPKGTDFERY 186
Db 127 NVDCCKDALLAKLESANVYKALILSDYKGTLDKQKMIQARKANVEVLIDPKGTDFERY 186
QY 187 RGATLLTPNLSEFAVGVCKTKEEIVERGMKLIADYELSALLVTRSEQMSLLQPKAP 246
Db 187 RGATLLTPNMSEFAVGVCKTKEEIVERGMKLIADYELSALLVTRSEQMSLLRNOEP 246
QY 247 LHPTQAEYVDVTGAGDTVIGVLAATLAAGNSLEACFPFANAAGVGVCKLGTSTVSP 306
Db 247 YHLPVAKVEFDTGAGDIVISVLAATLADGRSFEECYLANVAAGVGVCKLGTSTVSTV 306
QY 307 ELENAVGRADTGFVGMTEELKLAFAAARKGKVVMTNGVDFILHAGHVSYLANARKL 366
Db 307 ELENAVHARPETGFGIMSEALKDAVAQAKARGEKIVMTNGCFDILHPGHISYLENARKL 366
QY 367 GRLIVAVNSDASTKRLKGDSPVNPLEQRMVILGALEADVWVSEEDTPQRLIAGILP 426
Db 367 GRLIVAVNSDSSVKRLKGESRPINLENRMVAVLAGSVDMVLVFTEDTPQRLIGEILP 426
QY 427 DLLVKGDDYKPEELAGSKVWANGGEVLVLFNFDGCGSTTNIKKIQ 472
Db 427 DLLVKGDDYKPEELAGSKVWANGGEVLVLFNFDGCGSTTNIKKIK 472

RESULT 3
ID RFAE_HELPJ STANDARD; PRT; 463 AA.
AC O92KZ0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADP-heptose synthase (EC 2.7.-.-).
GN RFAE OR WAAE OR JHP0792.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
RA Tummino P., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.

```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE001509; AAD06368.1; -
DR HSSP; P27623; ICOZ
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfkB; 1.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR TIGRfams; TIGR00125; Cyt_tran_rel; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; FALSE_NEG.
DR PROSITE; PS00584; PFKB_KINASES_2; FALSE_NEG.
DR Lipopolysaccharide biosynthesis; Transferase; Complete proteome.
KW SEQUENCE 463 AA; 50926 MW; 770367403E771124 CRC64;
Query Match 34.2%; Score 819.5; DB 1; Length 463;
Best Local Similarity 41.0%; Pred. No. 6.5e-42;
Matches 193; Conservative 87; Mismatches 166; Indels 25; Gaps 9;
QY 13 VAVGVDMVDRLRYWGPTSRISPEAPVVKVNTIEERPGGAANVAMNIASLGANARLVGL 72
DB 4 ILVIGDLADYLYLWAGKSERLSPEAPVPLEVKKESKNGGAANVANNITSLKARVLCGV 63
QY 73 TGIDDAARALSKSLADVNVKDFVSV-PPTHPTITKLRLVLSRNQQLIRLDFEEFGVDP- 130
DB 64 VGDDLEGKHFSTLKTGRIDTSGVLIDKTRCTTLKTRIAAQNOQIVRVDEI----KQPL 119
QY 131 -----QPLHERINQALSSIGALVLSYAKGAL--ASVQOMQLARKAGVPLIDPKGTF 183
DB 120 NADLRKLLDTFAEKIQEIDGVILSDYKNGVLDLFDLTITLANKHKLILCDPKGKY 179
QY 184 ERYRGATLLTPNLSEFEAVGCKTEEEIVERGMKLIAD-YELSAALLVTRSGGSLLOP 242
DB 180 SKYSHASLITPNRALEQALHLKLDHANLSKALQILOETHYIAMPVLTSEQIAFLEK 239
QY 243 GKAPLHPTQAOEYVDVTGAGTVIGVLAATLAAGNSLSEACFFANAAAGVVGKLTGTST 302
DB 240 GEL-VNCPITAEVYVDVTGAGTVIASLTLSLESKSLKEACEFANAAAVVVGKMGSL 298
QY 303 VSPTELENVGRADTGFVWTEELKLVAAARKGEKVVMTNGVFDILHAGHVSILAN 362
DB 299 AS---LEEIALINQTHPKILPLEKL---LETLRNQOKIVFTNGCFDILHKGHASYLQK 352
QY 363 ARKGLDRILVAVNSDASTKRLKGDSPVNPLEQRMIVLGALEAVDWMVYSFEEDTPORLIA 422
DB 353 AKALGDLVGLNSDNSIKRLKGRPIVSEKDRAFILASLSCVDYVVVFGEDTPIKLIQ 412
QY 423 GILPDLLVKGDKYKPEIAGSKYEWANGVVLVLFNFDGCGCTTNIKKIQQ 473
DB 413 ALKPDILVKGADYLNKEYIGSE---LAKETRLIEFEGYSTSAIIKIKR 459
RESULT 4
Y060_BUCAI
ID Y060_BUCAI STANDARD; PRT; 315 AA.
AC P57168;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical sugar kinase BU060.
GN BU060.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-TOKYO 1998;
RX MEDLINE-20445173; PubMed-10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AP001118; BABI2783.1; -
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfkB; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; FALSE_NEG.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
KW SEQUENCE 315 AA; 35023 MW; 69B3707C601EFD25 CRC64;
Query Match 33.9%; Score 811; DB 1; Length 315;
Best Local Similarity 50.0%; Pred. No. 1.3e-41;
Matches 156; Conservative 63; Mismatches 93; Indels 0; Gaps 0;
QY 1 MKVTLPFERAGVWVGVDMVDRLRYWGPTSRISPEAPVVKVNTIEERPGGAANVAMNI 60
DB 1 MKKKLINFNSLVVGVGLIILDCYWSKNHYMSQLTPVPIVINKIKEOPGGAANVAKNI 60
QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLRLVLSRNQQLIRLD 120
DB 61 AEIGGYTKIVGFIGNEGILLKLMHDHIRDDLSISKNKNTITKIRLSEKKQLIRVD 120
QY 121 FEEGEGVDPOPLHERINQALSSIGALVLSYAKGALASVQOMQLARKAGVPLIDPKG 180
DB 121 FQEKYISKNKLLHOKIIDSISSFKILVLSYAKGTLANIQNIIDLAKKMSIPILDPKG 180
QY 181 TDFERYGATLLTPNLSEFEAVGCKTEEEIVERGMKLIADYELSAALLVTRSGGMSLL 240
DB 181 IDFKYSGASLLTPNLFEFEKIVGKCYRENEILRGINKLLSELQISALLVTRSKNGMTLF 240
QY 241 QPKAPLHPTQAOEYVDVTGAGTVIGVLAATLAAGNSLEACFFANAAAGVVGKLTGT 300
DB 241 QKEKPIHFPAAASKTASDVGTGADTVIALIAASLATGYSLEEACFYANIGASIVIQKIGT 300
QY 301 STVSPTELENV 312
DB 301 ETLNINELNSVL 312
RESULT 5
RFAE_HELPY
ID RFAE_HELPY STANDARD; PRT; 461 AA.
AC O25529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADP-heptose synthase (EC 2.7.-.-).
GN RFAE OR WAAE OR HP0858.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kurlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
```


RA Entian K.D., Errington J., Fabret C., Ferrazi E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.-J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinolis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roches E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB007638; BAA22760.1; -;
DR EMBL; Z99107; CAB12436.1; -;
DR HSP; G97VM2; 1DG5.
DR Subtilist; BG12796; ydJE.
DR InterPro; IPR002173; pfKB.
DR Pfam; PF00294; pfKB; 1.
DR PROSITE; PS00583; PFKB_KINASES.1; 1.
DR PROSITE; PS00584; PFKB_KINASES.2; 1.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 320 AA; 34256 MW; A54E09503953B7A CRC64;

Query Match 6.9%; Score 164.5; DB 1; Length 320;
Best Local Similarity 24.5%; Pred. No. 0.004;
Matches 80; Conservative 51; Mismatches 141; Indels 55; Gaps 12;

QY 13 VMVVGDMVLDVWYGYGPTSRISPEAPVPVVKVNTIERPGGA-ANVAMNIAISLGANARLVG 71
Db 6 VVCIGELLIDFF-----CTDVDLMGEGQFLKSAGGAPANVSAATKLGGDRAFSG 57
QY 72 LTGIDDAARALSKSLADVNVKDFSVPTHTTTKLRLVLSRQQLIR-LDFEEG----- 124
Db 58 KVGKDPFGYFLKRTLDVHVDTSM-L-VMDEKAPTTLAFVSLKQNGRDFVFNKGADALPT 116
QY 125 FEGVDQPLHERINQALSSIGALVLSYDAKALASVQMIQARAGVPLIDPKGTDFE 184
Db 117 LEDIDQEKINEAKILHFGSATALLSDPFCSAYL-----RLMSIAKONGQFISDFDNYRE-D 171
QY 185 RYRGATLLTPNLSEFAVVGK-----KTEIEIV-----ERGMKLADVLSAL 228
Db 172 LWRG-----RVSEFVSVAKKAIAVSDFKVSDDELEIISGVKDHKGVAIILHEIGANIV 225
QY 229 LVTRSGQMSLLQPGKALHPTQAEVYDVVTGAGDTVIGVLAATLAAGN----- 278
Db 226 AVTLGRSG-TLLSNGKDRIEIPTSIPVTSIDTSGAGDAFVGAALYQANTDQIOSVDADFV 284
QY 279 SLEEACFFANAAAGVVGKLGSTVSP 305
Db 285 KUREIVAFANKGALVCTKIGIDALP 311

RESULT 12
K6P2_ECOLI STANDARD; PRT; 309 AA.
ID K6P2_ECOLI
AC P06999; P78065; P78260;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase isozyme 2 (EC 2.7.1.11) (Phosphofructokinase-2).
GN PFKB OR B1723.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=84262485; PubMed=6235149;
RA Daidal F.;
RT "Nucleotide sequence of gene pfkb encoding the minor
RT phosphofructokinase of *Escherichia coli* K-12.";
RL Gene 28:337-342(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nasimondam H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [4]
RP SEQUENCE OF 1-38 FROM N.A.
RX MEDLINE=83294514; PubMed=6310120;
RA Daidal F.;
RT "Molecular cloning of the gene for phosphofructokinase-2 of
RT *Escherichia coli* and the nature of a mutation, pfk1, causing a high
RT level of the enzyme.";
RL J. Mol. Biol. 168:285-305(1983).
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -!- ENZYME REGULATION: PFK-2 IS SENSITIVE TO INHIBITION BY FRUCTOSE
CC 1,6-DIPHOSPHATE.
CC -!- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMODIMER.
CC -!- MISCELLANEOUS: ONLY 10% OF THE ACTIVITY PRESENT IN THE WILD-TYPE
CC STRAIN IS PHOSPHOFRUCTOKINASE-2.
CC -!- MISCELLANEOUS: THIS ENZYME IS NOT TO BE CONFUSED WITH
CC 6-PHOSPHOFRUCTO-2-KINASE (EC 2.7.1.105), WHICH IS ALSO CALLED
CC PHOSPHOFRUCTOKINASE 2.
CC -!- SIMILARITY: PHOSPHOFRUCTOKINASE-1 SHOWS NO HOMOLOGY TO
CC PHOSPHOFRUCTOKINASE-2, THE MINOR PHOSPHOFRUCTOKINASE FOUND IN
CC *E. COLI*. ALTHOUGH THE REACTION CATALYSED IS THE SAME, THE TWO
CC ENZYMES HAVE A DIFFERENT EVOLUTIONARY ORIGIN.
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; K02500; AAA24321.1; -
 DR EMBL; AE000267; AAC74793.1; -
 DR EMBL; D90814; BAA15500.1; ALT_INIT.
 DR EMBL; D90815; BAA15506.1; ALT_INIT.
 DR EMBL; K00128; AAA24320.1; -
 DR PIR; A24950; KIECFB.
 DR SWISS-2DPAGE; P06999; COLI.
 DR ECO2DBASE; E036.6; 6TH EDITION.
 DR ECoGene; EG10700; pfkb.
 DR InterPro; IPR002173; pfkb.
 DR Pfam; PF00294; pfkb; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; 1.
 DR PROSITE; PS00584; PFKB_KINASES_2; 1.
 KW Transferase; Kinase; Glycolysis; Complete proteome.
 FT CONFLICT 26 38 GKLRCAPVPEPG -> ENCAVPHRCNP (IN REF. 1
 FT AND 4);
 FT CONFLICT 155 171 AAKQGRICIVDSGGEA -> LRKNKGSAASTVVGOG
 FT (IN REF. 1);
 FT CONFLICT 245 246 PV -> AL (IN REF. 1);
 FT CONFLICT 257 258 SM -> RL (IN REF. 1);
 FT SEQUENCE 309 AA; 32456 MW; A93BBEOD5801309 CRC64;

Query Match 6.6%; Score 158; DB 1; Length 309;
 Best Local Similarity 25.4%; Pred. No. 0.0092;
 Matches 80; Conservative 37; Mismatches 122; Indels 76; Gaps 14;

QY 29 TSRISE-----APVPVKVNTTEPRGGAANYAMIASLGANARLV-----GLTGIDDA 78
 DB 19 TPQIYPEGKLRCAPV-----FEPGGGINVARIAIHGGSATAIFPAGGATG----- 66
 QY 79 ARALSKSLADVNVKCDFFVSPHTITKRLVLSRNOQLRLDFF-----EGFEGVDP----- 130
 DB 67 -EHLVSLADENV-----PVATVEAKDWTQRNLHVHVEASGEQYRFVMPGAALN 114
 QY 131 ----QPLHERINQALSSIGALVLSYAKGA-LASVQOMIQARAGVVPVLIDPKGTDFER 185
 DB 115 EDEFROLEEQVLETSAGAILVIGSLPPGVKLEKLTQLISAQKQRCIVDSSG---EA 171
 QY 186 YRGA-----TLTPNLSEFAVVGK-----CKTEEEIVERG--MKLADYELSALL 229
 DB 172 LSAALATGNIELVPKQKELSAVYNRELTPDDYKRAQELVNSGKAKRVVSLGPQCAL 231
 QY 230 VTRSEQMSLQPGKAPLHMPTOAQEYVDYVTGAGDTVIGVLAATLAAGNSLEEAACFTANA 289
 DB 232 GVDSENCIQVVP-----PVKSQ-----STVGAGDSMVGAMTLKLAENASLEENVRFGVA 281
 QY 290 AAGVVVGKLGSTVYS 304
 DB 282 AGSAATLNQGRILCS 296

RESULT 13

SCRK_VIBAL STANDARD; PRT; 307 AA.
 ID SCRK_VIBAL
 AC P22824;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Fructokinase (EC 2.7.1.4).
 GN SCRK.
 OS Vibrio alginolyticus.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=663;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91071601; PubMed=2174811;
 RA Blatch G.L., Scholle R.R., Woods D.R.;

RT "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
 RT uptake-encoding region.";
 RL Gene 95:17-23(1990).
 CC -1- CATALYTIC ACTIVITY: ATP + D-fructose = ADP + D-fructose 6-
 CC phosphate.
 CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M76768; AAA27556.1; -
 DR PIR; JQ0782; JQ0782.
 DR HSSP; Q9TWM2; .DGY.
 DR InterPro; IPR002173; pfkb.
 DR Pfam; PF00294; pfkb; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; 1.
 DR PROSITE; PS00584; PFKB_KINASES_2; 1.
 KW Transferase; Kinase.
 SQ SEQUENCE 307 AA; 33045 MW; AF9C96CBB781CGEF CR364;

Query Match 6.4%; Score 153; DB 1; Length 307;
 Best Local Similarity 23.5%; Pred. No. 0.018;
 Matches 78; Conservative 53; Mismatches 131; Indels 70; Gaps 12;

QY 13 VMVYGDMRLRYWYGTSPRSISPEAPVPVKVNTTEPRGGA-ANVAMNIASLGANARLVG 71
 DB 4 VWTGDAVVD-----LIPESETSLKC-----PGGAPANVAIAIARLSCKSAFFG 48
 QY 72 LTGIDDAARALSKSLADVNVKCDFFVSPHTITKRLVLSRNOQLRLDFFEGFEGVDP 130
 DB 49 RVGDDPFGRTMQSLDQEGVCTEELIKDPEQRTST-----VVVDLDQGERSFT 97
 QY 131 QPLHERINQALS--SIGALVLSDYA-----KGLASVQOMIQARAGVVPVLIDPK 179
 DB 98 FMVPSADQMSVEDMGNFQGDMLHVCSISLANEPSRSSTFEAIKKAAGGISEDPN 157
 QY 180 GTDFERYGATLLTPNLSEFAVVGKCK-----TEEEI-----VERGMKLIADY 223
 DB 158 LRD-----EVNQDQSEIQAVYKAVAMADVVYKFSSEELFETDSTMAQGLQIAAM 209
 QY 224 ELSALLVTRSGNSLLQPGKAPLHMPTOAQEYVDYVTGAGDTVIGVLAATLAAGNS---- 279
 DB 210 NIALVLVTQAGKGVNRFESQSEL-ITGVVSPIDTTGAGDAFVGGLLACLRRHADWKNH 268
 QY 280 --LEEACFFANAAAGVVVGKLGSTVSPIDLE 309
 DB 269 PVVSSAIOWANGCGALATTQKGAMTALPTQTE 300

RESULT 14

RBSK_HUMAN STANDARD; PRT; 322 AA.
 ID RBSK_HUMAN
 AC Q9H477;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ribokinase (EC 2.7.1.15).
 GN RBSK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wightman P.J.;
 RL Thesis (2000), University of Edinburgh, U.K.
 RN [2]
 RP SEQUENCE FROM N.A.

```
RC TISSUE=Lung;
RA Submitted R.;
RL Submittal (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-ribose = ADP + D-ribose 5-phosphate.
CC -!- PATHWAY: Ribose metabolism; first step.
CC -!- SIMILARITY: BELONGS TO THE PKFB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ404857; CAC12877.1; -
DR EMBL; BC017425; AAI17425.1; -
DR HSSP; P05054; IRK2.
DR InterPro; IPR002173; PfkB.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfkB; 1.
DR PRINTS; PR00390; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; FALSE_NEG.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR Transferase; Kinase.
KW TRANSFERASE; KINASE.
SQ SEQUENCE 322 AA; 34143 MW; 50D0E7161F33E94B CRC64;

Query Match 6.4%; Score 153; DB 1; Length 322;
Best Local Similarity 25.5%; Pred. No. 0.019;
Matches 83; Conservative 48; Mismatches 140; Indels 54; Gaps 13;

QY 9 ERAGVAVGVDMVLDVRYGPTSRISPEAPVPVVKVNTIEERP-----GGAANVAMNIA 62
DB 14 EVAAVVVGSCMTD-LVSLTSRLPTG-----ETIGHKFFIGFGKGANQCVQAA 64
QY 63 LGANARLVGLTGIDDAARALSKLADVNVKDFSVVPHT-----ITKLVLSRNQOL 116
DB 65 LGAMTSWVKVGKDSFGNDYIENLQNDISTEF----TYQTKDAATGASIIVNEGONI 120
QY 117 IRLDFEEGEGVDPQPLHERINQALSSIG-ALVLSYAKGALASVQOMIOLARKAGVPVL 175
DB 121 IVI-----VAGANLLNTEEDRAANVTSRAKVVWVCOLEITPATSLTAMRRSGVKT 175
QY 176 IDP----KGTDFERYRGATLLPNLSEFAV-----VGKCKTEE-----IVERGMKLIADY 223
DB 176 FNPAPAIADLDQFYTLSDVFCNNESEAEILGLTVGSAADAGEAALVLLKRGQCVW--- 232
QY 224 ELSALLVTRSEOGMSLL-OPGKAPLHMTQAEVYDVDTGAGTVIGVLAATLA--AGNSL 280
DB 233 -----IITLGAEGCVVLSQTEPEPKHIPTKVKAVDTTGTAGDSFVGALAFYLAYYPNLSL 287
QY 281 ERACFFANAAAGVYVKGKLGTSVSP 305
DB 288 EDMNLSNFIAVSVQAACTQSSYP 312

RESULT 15
SCRK_SALTY STANDARD; PRT; 307 AA..
AC P26984;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Fructokinase (EC 2.7.1.4).
GN SCRK.
OS Salmonella typhimurium.
OG Plasmid pUR400.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92236409; PubMed=1809835;
```

```
RA Aulkemeyer P., Ebner R., Heilenmann G., Jahreis K., Schmid K.,
RA Wrieden S., Lengeler J.W.;
RT "Molecular analysis of two fructokinases involved in sucrose
RT metabolism of enteric bacteria.";
RL Mol. Microbiol. 5:2913-2922(1991).
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose = ADP + D-fructose 6-
CC phosphate.
CC -!- SIMILARITY: BELONGS TO THE PKFB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61005; CAA43323.1; -
DR PIR; S16044; S16044.
DR PIR; S18524; S18524.
DR HSSP; Q9TVW2; IDGY.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00294; pfkB; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR Transferase; Kinase; Plasmid.
KW TRANSFERASE; KINASE.
SQ SEQUENCE 307 AA; 32916 MW; E01CB770CE20B329 CRC64;

Query Match 6.4%; Score 152.5; DB 1; Length 307;
Best Local Similarity 23.8%; Pred. No. 0.019;
Matches 81; Conservative 48; Mismatches 126; Indels 85; Gaps 14;

QY 11 AGVAVGVDMVLDVRYGPTSRISPEAPVPVVKVNTIEERP-----GGAANVAMNIA 69
DB 3 AKVAVVVGDAVD-----LLPESE-----GRLLQCPGAPANVAVGVLGGNSGF 47
QY 70 VOLTGIDDAARALSKLADVNVKDFSVV-PHTPTITKLRLV-----SRNQ 114
DB 48 IGAVGDPFGRYMRHTLQOEQVDVSHMYLDDQHRTSTVVDLDDOGERFTFMWRPSADL 107
QY 115 OLIRLDFEEGEGVDPQPLHERINQALSSIGALVLSYAKGALASVQOMIOLARKAGVPV 174
DB 108 FLVEEDLPQFAAG---QWLH-----VCSIALS--AEPSTTFAMESIRSAGGRV 153
QY 175 LIDPKGTDFERYRGATLLPNLSEFAVVGK-----KTEEEIV-----E 214
DB 154 SFDPN-----IRPDLWQDQALLACLDRALHMANVVKUSEELVFISSSNDIA 201
QY 215 RGMKLIAD-YELSALLVTRSEOGMSLLQPGKAPLHMTQAEVYDVDTGAGTVIGVLAAT 273
DB 202 YGIASVTERYQPELLLVTRGKAGVLAAFQOKF-THFNARPVASVDVDTGAGDAFVAGLLAS 260
QY 274 LAAG-----NSLEACFFANAAAGVYVKGKLGTSVSP 307
DB 261 LAANGMPTDMTALPTLLLAQTGCGALATTAKGANTALPYQ 300

Search completed: November 24, 2002, 23:16:40
Job time : 40 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2002, 23:11:44 : Search time 95 Seconds

(without alignments)
1034.573 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLPFERAGVMVGVDM.....FEDGCGTIIKIQDKKG 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2385	99.7	477	16 Q8XBM4	Q8xbm4 escherichia
2	2276	95.1	477	2 Q9AJ74	Q9aj74 salmonella
3	2265	94.7	477	16 Q8XEW9	Q8xew9 salmonella
4	2250	94.0	477	2 Q9RFY8	Q9rfy8 salmonella
5	2225	93.0	473	2 Q9RFY7	Q9rfy7 salmonella
6	2076	86.8	476	16 Q8Z160	Q8zi60 yersinia pe
7	1701	71.1	476	16 Q9CME6	Q9cm6 pasteurilla
8	1350.5	56.4	474	16 Q9HUG9	Q9hug9 pseudomonas
9	1122	46.9	342	2 Q48046	Q48046 haemophilus
10	904.5	37.8	483	16 Q9A2C5	Q9a2c5 campylobacter
11	863	36.1	461	16 Q9PNE5	Q9pne5 campylobacter
12	863	36.1	496	16 Q98154	Q98i54 rhizobium l
13	753.5	31.5	313	2 Q9X518	Q9x518 neisseria m
14	753.5	31.5	323	16 Q9K004	Q9k004 neisseria m
15	745.5	31.2	323	16 Q9JRJ4	Q9jrj4 neisseria m
16	721.5	30.2	319	16 Q8Y0X9	Q8y0x9 ralstonia s

17	678.5	28.4	307	2 Q9WXX7	Q9wx7 burkholderi
18	610.5	25.5	498	2 Q9FB11	Q9fb11 streptomyce
19	594	24.8	463	16 Q9Z5B5	Q9z5b5 streptomyce
20	584.5	24.4	323	16 Q8R6E8	Q8r6e8 fusobacteri
21	564.5	23.6	315	16 Q66836	Q66836 aquifex aeo
22	407.5	17.0	160	16 Q8RF00	Q8rf00 fusobacteri
23	391.5	16.4	157	16 Q66572	Q66572 aquifex aeo
24	368.5	15.4	150	2 Q51060	Q51060 neisseria g
25	348.5	14.6	168	16 Q9JXF0	Q9jxf0 neisseria m
26	340.5	14.2	168	16 Q9JW18	Q9jw18 neisseria m
27	314	13.1	166	16 Q8Y2M6	Q8y2m6 ralstonia s
28	311.5	13.0	164	2 Q43999	Q43999 alcaligenes
29	221.5	9.3	349	16 Q8RBP6	Q8rbp6 thermoanaer
30	201.5	8.4	319	16 Q9KBR8	Q9kbr8 bacillus ha
31	197	8.2	308	16 Q912F4	Q912f4 pseudomonas
32	194	8.1	299	16 Q9X055	Q9x055 thermotoga
33	192.5	8.0	298	16 Q9L7R2	Q9l7r2 salmonella
34	190.5	8.0	306	16 Q8RD45	Q8rd45 thermoanaer
35	190	7.9	323	11 Q8R109	Q8r1q9 mus musculus
36	189.5	7.9	148	17 Q9UZ37	Q9uz37 pyrococcus
37	188	7.9	307	16 Q8YW32	Q8yw32 anabaena sp
38	185	7.7	309	16 Q982U3	Q982u3 rhizobium l
39	184.5	7.7	303	16 Q9A9M6	Q9a9m6 caulobacter
40	183.5	7.7	319	16 Q8Z2W9	Q8z2w9 salmonella
41	183.5	7.7	323	10 Q944F4	Q944f4 oryza sativ
42	182.5	7.6	148	17 Q8U1T9	Q8u1t9 pyrococcus
43	181.5	7.6	148	17 Q58466	Q58466 pyrococcus
44	181.5	7.6	378	10 Q9SHH5	Q9shh5 arabidopsis
45	181	7.6	315	16 Q8Y0N0	Q8y0n0 ralstonia s

ALIGNMENTS

RESULT 1

Q8XBM4	PRELIMINARY;	PRT;	477 AA.
ID	Q8XBM4		
AC	Q8XBM4		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Putative kinase (ADP-heptose synthase).		
GN	Z4405 OR ECS3935.		
OS	Escherichia coli O157:H7.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=83334;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;		
RX	MEDLINE=21074935; PubMed=11206551;		
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,		
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,		
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,		
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,		
RA	Welch R.A., Blattner F.R.;		
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"		
RL	Nature 409:529-533(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / RMD 0509952;		
RX	MEDLINE=21156631; PubMed=11258796;		
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,		
RA	Han C.-G., Ohkubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,		
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,		
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;		
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli		
RT	O157:H7 and genomic comparison with a laboratory strain K-12.;"		
RL	DNA Res. 8:11-22(2001).		
DR	EMBL; AF005534; AAG58186.1; -		
DR	EMBL; AF002564; BAB37358.1; -		
DR	InterPro; IPR004820; Cytidylyltransf.		

DR InterPro; IPR004821; Cyt_tran_rel.
 DR InterPro; IPR002173; pfkB.
 DR Pfam; PF01467; Cytidylyltransf; 1.
 DR Pfam; PF00294; pfkB; 1.
 DR TIGRFAMs; TIGR00125; Cyt_tran_rel; 1.
 DR PROSITE; PS00583; pfkB_KINASES_1; UNKNOWN_1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 477 AA; 51064 MW; 51064 MW; BB877FEF6636E67C CRC64;

Query Match 99.78; Score 2385; DB 16; Length 477;
 Best Local Similarity 99.68; Pred. No. 2.1e-142;
 Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVGVGDVMDLDRYWGTPSRISPEAPVPVVKVNTIIEERPGGAANVAMNI 60
 Db 1 MKVTLPEERAGVGVGDVMDLDRYWGTPSRISPEAPVPVVKVNTIIEERPGGAANVAMNI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTTTKLRVLSRNQOLIRLD 120
 Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTTTKLRVLSRNQOLIRLD 120

Qy 121 FEEFEGVDPOPLHERINQALSSIGALVSDYAKGALASVOOMIOLARKAGVPLIDPKG 180
 Db 121 FEEFEGVDPOPLHERINQALSSIGALVSDYAKGALASVOOMIOLARKAGVPLIDPKG 180

Qy 181 TDFERYRGATLLTPNLSEFAVVGKCKTEEIVVERGMKLIADYELSAALLVTRSEQGSLL 240
 Db 181 TDFERYRGATLLTPNLSEFAVVGKCKTEEIVVERGMKLIADYELSAALLVTRSEQGSLL 240

Qy 241 QPKAPLHMPQAOEVDVDTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVVVKLG 300
 Db 241 QPKAPLHMPQAOEVDVDTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVVVKLG 300

Qy 301 STVSPIELENVAVRGRADTGFVMTTEELKLAFAAARKRGEKVMTNGVFDILHAGHSYL 360
 Db 301 STVSPIELENVAVRGRADTGFVMTTEELKLAFAAARKRGEKVMTNGVFDILHAGHSYL 360

Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALFAVDWVVSFEEDTPQRL 420
 Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALFAVDWVVSFEEDTPQRL 420

Qy 421 IAGILPOLLVKGDDYKPEETAGSKEVWANGGEVLVLPFEDGCSSTNIKKIQDQKK 477
 Db 421 IAGILPOLLVKGDDYKPEETAGSKEVWANGGEVLVLPFEDGCSSTNIKKIQDQKK 477

RESULT 2
 Q9AJ74
 ID Q9AJ74 PRELIMINARY; PRT; 477 AA.
 AC Q9AJ74
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE ADP-heptose synthase (Fragment).
 GN RFAE.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jin U.-H., Chung T.-W., Kim C.-H.;
 RT "ADP-heptose synthase (rfae) gene of salmonella typhimurium."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF155126; AAK20933.1; .
 DR HSP; P27623; 1COZ.
 DR InterPro; IPR004820; Cytidylyltransf.
 DR InterPro; IPR004821; Cyt_tran_rel.
 DR Pfam; PF01467; Cytidylyltransf; 1.
 DR Pfam; PF00294; pfkB; 1.
 DR TIGRFAMs; TIGR00125; Cyt_tran_rel; 1.
 DR PROSITE; PS00583; pfkB_KINASES_1; UNKNOWN_1.

FT NON_TER 477 477
 SQ SEQUENCE 477 AA; 51169 MW; 269475F3FF9EB9EF CRC64;

Query Match 95.18; Score 2276; DB 2; Length 477;
 Best Local Similarity 93.98; Pred. No. 1.6e-135;
 Matches 447; Conservative 19; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVGVGDVMDLDRYWGTPSRISPEAPVPVVKVNTIIEERPGGAANVAMNI 60
 Db 1 MKVTLPEERAGVGVGDVMDLDRYWGTPSRISPEAPVPVVKVNTIIEERPGGAANVAMNI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTTTKLRVLSRNQOLIRLD 120
 Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTTTKLRVLSRNQOLIRLD 120

Qy 121 FEEFEGVDPOPLHERINQALSSIGALVSDYAKGALASVOOMIOLARKAGVPLIDPKG 180
 Db 121 FEEFEGVDPOPLHERINQALSSIGALVSDYAKGALASVOOMIOLARKAGVPLIDPKG 180

Qy 181 TDFERYRGATLLTPNLSEFAVVGKCKTEEIVVERGMKLIADYELSAALLVTRSEQGSLL 240
 Db 181 TDFERYRGATLLTPNLSEFAVVGKCKSEDELVERGMKLIADYELSAALLVTRSEQGSLL 240

Qy 241 QPKAPLHMPQAOEVDVDTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVVVKLG 300
 Db 241 QPKAPLHMPQAOEVDVDTGAGDTVIGVLAATLAAGNTLEEACFFANAAAGVVVVKLG 300

Qy 301 STVSPIELENVAVRGRADTGFVMTTEELKLAFAAARKRGEKVMTNGVFDILHAGHSYL 360
 Db 301 STVSPIELENVAVRGRADTGFVMTTEELKLAFAAARKRGEKVMTNGVFDILHAGHSYL 360

Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALFAVDWVVSFEEDTPQRL 420
 Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALFAVDWVVSFEEDTPQRL 420

Qy 421 IAGILPOLLVKGDDYKPEETAGSKEVWANGGEVLVLPFEDGCSSTNIKKIQDQKK 476
 Db 421 IAGILPOLLVKGDDYKPEETAGSKEVWANGGEVLVLPFEDGCSSTNIKKIQDQKK 476

RESULT 3
 Q8XEW9
 ID Q8XEW9 PRELIMINARY; PRT; 477 AA.
 AC Q8XEW9
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Bifunctional, putative sugar nucleotide transferase domain of
 DE ADP-L-glycero-D-manno-heptose synthase (EC 2.7.-.-) (ADP-heptose
 DE synthase).
 GN RFAE OR STM3200 OR STY3379.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2."
 RL Nature 413:852-856(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18.
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,


```
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR ENBL; AF008847; AAL22074.1; -.
DR EMBL; ALG27278; CAB07725.1; -.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR Pfam; PF00294; pfkB; 1.
DR TIGRFAMS; TIGR00125; cyt_tran_rel; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
KW Transferase; Complete proteome.
SQ SEQUENCE 477 AA; 51124 MW; EAFFG6B1DEE80568C CRC64;

Query Match 94.78; Score 2265; DB 16; Length 477;
Best Local Similarity 93.58; Pred. No. 7.9e-135;
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVVMVGDVMDLDRWYGTSTRISPEAPVPVVKVNTIERPGGAANVANNI 60
Db 1 MKVNLPAFERAGVVMVGDVMDLDRWYGTSTRISPEAPVPVVKVNTIEERPGGAANVANNI 60
Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTTTIKLRLVLSRNOQLRLD 120
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTTTIKLRLVLSRNOQLRLD 120
Qy 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOQMLARKAGVPVLIDPKG 180
Db 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOQMLARKAGVPVLIDPKG 180
Qy 181 TDFERYGATLLTPNLSEFAVVGKCTEEIEVERGMKLIADYLSALLVTRSEOGMSLL 240
Db 181 TDFERYGATLLTPNLSEFAVVGKCTEEIEVERGMKLIADYLSALLVTRSEOGMSLL 240
Qy 241 QPKAPLHMPQAOEVYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300
Db 241 QPNKAPLHMPQAOEVYDVTGAGDTVIGVLAATLAAGNTLEEACVFANAAAGVVGKLG 300
Qy 301 STVSPIELENVAVRGADTGFVGMTEELKLAFAAARKRGEKVVMTNGVDFILHAGHVSYL 360
Db 301 STVSPIELENVAVRGADTGFVGMTEELKLAFAAARKRGEKVVMTNGVDFILHAGHVSYL 360
Qy 361 ANARKLGDRLLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALVAVDWMVVSFEEDTPORL 420
Db 361 ANARKLGDRLLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALVAVDWMVVSFEEDTPORL 420
Qy 421 IAGILPDLVLKGGDYKPEIAGSEKVGWANGGEVLVLPNEDGCSSTNTIHKIQOQKK 476
Db 421 IAGILPDLVLKGGDYKPEIAGSEKVGWANGGEVLVLPNEDGCSSTNTIHKIQOQKK 476

RESULT 4
Q9RFY8 PRELIMINARY; PRT; 477 AA.
ID AC
Q9RFY8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RfaE.
GN RFAE.
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=28901;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=SL1027;
RX MEDLINE=20096694; PubMed=10629197;
RA Valvano M.A., Marolda C.L., Bittner M., Glaskin-Clay M., Simon T.L.,
RA Klena J.D.;
RT "The rfaE gene from Escherichia coli encodes a bifunctional protein
RT involved in the biosynthesis of the lipopolysaccharide core precursor
RT ADP-L-glycero-D-manno-heptose."
RL J. Bacteriol. 182:488-497(2000).
DR ENBL; AF163661; AAD49846.1; -.
DR HSSP; P27623; ICOZ.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR Pfam; PF00294; pfkB; 1.
DR TIGRFAMS; TIGR00125; cyt_tran_rel; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
SQ SEQUENCE 477 AA; 51119 MW; 9215BA18BD655FF7 CRC64;

Query Match 94.08; Score 2250; DB 2; Length 477;
Best Local Similarity 92.98; Pred. No. 7e-134;
Matches 442; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVVMVGDVMDLDRWYGTSTRISPEAPVPVVKVNTIEERPGGAANVANNI 60
Db 1 MRVNLPAFERAGVVMVGDVMDLDRWYGTSTRISPEAPVPVVKVNTIEERPGGAANVANNI 60
Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTTTIKLRLVLSRNOQLRLD 120
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNKCDVSVPTHTTTIKLRLVLSRNOQLRLD 120
Qy 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOQMLARKAGVPVLIDPKG 180
Db 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOQMLARKAGVPVLIDPKG 180
Qy 181 TDFERYGATLLTPNLSEFAVVGKCTEEIEVERGMKLIADYLSALLVTRSEOGMSLL 240
Db 181 TDFERYGATLLTPNLSEFAVVGKCTEEIEVERGMKLIADYLSALLVTRSEOGMSLL 240
Qy 241 QPKAPLHMPQAOEVYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300
Db 241 QPNKAPLHMPQAOEVYDVTGAGDTVIGVLAATLAAGNTLEEACVFANAAAGVVGKLG 300
Qy 301 STVSPIELENVAVRGADTGFVGMTEELKLAFAAARKRGEKVVMTNGVDFILHAGHVSYL 360
Db 301 STVSPIELENVAVRGADTGFVGMTEELKLAFAAARKRGEKVVMTNGVDFILHAGHVSYL 360
Qy 361 ANARKLGDRLLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALVAVDWMVVSFEEDTPORL 420
Db 361 ANARKLGDRLLIVAVNSDASTKRLKDSRPVNPLEQRMIVLGALVAVDWMVVSFEEDTPORL 420
Qy 421 IAGILPDLVLKGGDYKPEIAGSEKVGWANGGEVLVLPNEDGCSSTNTIHKIQOQKK 476
Db 421 IAGILPDLVLKGGDYKPEIAGSEKVGWANGGEVLVLPNEDGCSSTNTIHKIQOQKK 476

RESULT 5
Q9RFY7 PRELIMINARY; PRT; 473 AA.
ID AC
Q9RFY7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RfaE mutant.
GN RFAE.
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=28901;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20096694; PubMed=10629197;
```

RA Valvano M.A., Marolda C.L., Bittner M., Glaskin-Clay M., Simon T.L.,
 RA Klena J.D.;
 RT "rfaE gene from Escherichia coli encodes a bifunctional protein
 RT involved in the biosynthesis of the lipopolysaccharide core precursor
 RT ADP-L-glycero-D-manno-heptose";
 RL J. Bacteriol. 182:488-497(2000).
 DR EMBL; AF163662; AAD49847.1; -.
 DR HSP; P27623; 1C0Z.
 DR InterPro; IPR004820; Cytidylyltransf.
 DR InterPro; IPR004821; Cyt_tran_rel.
 DR InterPro; IPR002173; PfKB.
 DR Pfam; PF01467; Cytidylyltransf; 1.
 DR Pfam; PF00294; PfKB; 1.
 DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
 FT VARIANT 236 236 E -> G.
 SQ SEQUENCE 473 AA; 50868 MW; 829C1A377A508E6 CRC64;

Query Match 93.0%; Score 2225; DB 2; Length 473;
 Best Local Similarity 92.2%; Pred. No. 2.6e-132;
 Matches 439; Conservative 20; Mismatches 13; Indels 4; Gaps 1;
 QY 1 MKVTLPEFERAGVMVGVDMVDRIYWGPTSRISPEAPVPVVKVNTIEERPGGAANYAMNI 60
 DB 1 MRVNLPAFERAGVMVGVDMVDRIYWGPTSRISPEAPVPVVKVNTIEERPGGAANYAMNI 60
 QY 61 ASLGANARLVGLTGIDDAARALSKSLADYVNVKCDVSVTPHTTITKLRLVSRNQQLRLD 120
 DB 61 ASLGANARLVGLTGIDDAARALSKSLADYVNVKCDVSVTPHTTITKLRLVSRNQQLRLD 120
 QY 121 FEBGFEVDPQPLHERINQALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
 DB 121 FEBGFEVDPQPLHERINQALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
 QY 181 TDFERYRGATLLTPNLSEFEAVVGVCKTEEEIVERGMKLIADYELSAVLTRSEQMSLL 240
 DB 181 TDFERYRGATLLTPNLSEFEAVVGVCKTEEEIVERGMKLIADYELSAVLTRSEQMSLL 240
 QY 241 QPKAPLHPMPTQAEVYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAGVVVVKLG 300
 DB 241 QPKAPLHPMPTQAEVYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAGVVVVKLG 300
 QY 301 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
 DB 297 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 356
 QY 361 ANARKLGDRLLIYAVNSDASTKRLKGSRPVNPVLEQRMIVLGALEAVDWWVSPFEEDTPQRL 420
 DB 357 ANARKLGDRLLIYAVNSDASTKRLKGSRPVNPVLEQRMIVLGALEAVDWWVSPFEEDTPQRL 416
 QY 421 IAGILPDLLVKGDDYKPEETAGSKVEWANGGEVVLNFDGCGSTTNIKKIQODK 476
 DB 417 IAGILPDLLVKGDDYKPEETAGSKVEWANGGEVVLNFDGCGSTTNIKKIQOTSE 472

RESULT 6
 Q82160 PRELIMINARY; PRT; 476 AA.
 ID Q82160
 AC Q82160
 DT 01-MAR-2002 (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
 DE ADP-heptose synthase (EC 2.7.-.-).
 GN RFAE OR YP00654.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OC NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,

RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Hoiroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Parrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";
 RL Nature 413:523-527(2001).
 DR EMBL; AJ414144; CAC89508.1; -.
 DR InterPro; IPR004820; Cytidylyltransf.
 DR InterPro; IPR004821; Cyt_tran_rel.
 DR InterPro; IPR002173; PfKB.
 DR Pfam; PF01467; Cytidylyltransf; 1.
 DR Pfam; PF00294; PfKB; 1.
 DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
 KW Transference; Complete proteome.
 SQ SEQUENCE 476 AA; 51218 MW; E68FF382DC892636 CRC64;
 Query Match 86.8%; Score 2076; DB 16; Length 476;
 Best Local Similarity 86.1%; Pred. No. 6.6e-123;
 Matches 409; Conservative 32; Mismatches 34; Indels 0; Gaps 0;
 QY 1 MKVTLPEFERAGVMVGVDMVDRIYWGPTSRISPEAPVPVVKVNTIEERPGGAANYAMNI 60
 DB 1 MKVTLPEFERAGVMVGVDMVDRIYWGPTSRISPEAPVPVVKVNTIEERPGGAANYAMNI 60
 QY 61 ASLGANARLVGLTGIDDAARALSKSLADYVNVKCDVSVTPHTTITKLRLVSRNQQLRLD 120
 DB 61 ASLGANARLVGLTGIDDAARALSKSLADYVNVKCDVSVTPHTTITKLRLVSRNQQLRLD 120
 QY 121 FEBGFEVDPQPLHERINQALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
 DB 121 FEBGFEVDPQPLHERINQALSSIGALVLSYAKGALASVQOMIOLARKAGVPVLIDPKG 180
 QY 181 TDFERYRGATLLTPNLSEFEAVVGVCKTEEEIVERGMKLIADYELSAVLTRSEQMSLL 240
 DB 181 TDFERYRGATLLTPNLSEFEAVVGVCKTEEEIVERGMKLIADYELSAVLTRSEQMSLL 240
 QY 241 QPKAPLHPMPTQAEVYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAGVVVVKLG 300
 DB 241 QPKAPLHPMPTQAEVYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAGVVVVKLG 300
 QY 301 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
 DB 301 STVSPTELENAVGRADTGFVGMTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
 QY 361 ANARKLGDRLLIYAVNSDASTKRLKGSRPVNPVLEQRMIVLGALEAVDWWVSPFEEDTPQRL 420
 DB 361 ANARKLGDRLLIYAVNSDASTKRLKGSRPVNPVLEQRMIVLGALEAVDWWVSPFEEDTPQRL 420
 QY 421 IAGILPDLLVKGDDYKPEETAGSKVEWANGGEVVLNFDGCGSTTNIKKIQODK 475
 DB 421 IADILPDLLVKGDDYKPEETAGSKVEWANGGEVVLNFDGCGSTTNIKKIQKNGR 475

RESULT 7
 Q9CME6 PRELIMINARY; PRT; 476 AA.
 ID Q9CME6
 AC Q9CME6;
 DT 01-JUN-2001 (TremBLrel. 17, Created)
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
 DE RfaE.
 GN RFAE OR PM0884.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OC NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;

RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT	*Complete genomic sequence of <i>Pasteurella multocida</i> Pm70.*;
RL	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR	ENBL; AE006127; AAK02968.1; .
DR	HSP; P27623; 1COZ.
DR	InterPro: IPR004820; Cytidylyltransf.
DR	InterPro: IPR004821; Cyt_tran_rel.
DR	InterPro: IPR002173; pfkB.
DR	Pfam: PF01467; Cytidylyltransf; 1.
DR	Pfam: PF00294; pfkB; 1.
DR	TIGRFAMS: TIGR00123; cyt_tran_rel; 1.
DR	Complete proteome.
KW	SEQUENCE 476 AA; 51884 MW; C4137ECC7B2F02D1 CR664;

Query Match	71.1%	Score	1701;	DB	16;	Length	476;
Best Local Similarity	69.7%;	Pred.	No. 2.9e-9;				
Matches	327;	Conservative	65;	Mismatches	77;	Indels	0;
Gaps	0;						
Qy	7	EFERAGVVMVGDVMDRYWYGTSGRIPEAPVPVVKVNTIETERPAGGAANVAMNIAAS	126				
Db	7	EFKQAKVLVLGDVMDRYWFGATNRISPEAPVPVVDNEERAGGAANVAMNIAALNVP	66				
Qy	67	ARLVGLTGIDDAARALSKSLADVNVKDFVSPVTHPTITKLRVLSRNOQLRLDFE	126				
Db	67	VKLGGLGRDETGALESLLAKSRIEDCFVLQHTHTPTITKLRILSRHQQLRLDFE	126				
Qy	127	GVDPQPLHERINQALSSITGALVLDYAKGALASVQOIMIQARKAGVPLVIDPKGTDFERY	186				
Db	127	NVTSDALQLKLDLSALQVGVLDSDYKGTGLNVQVKMLQLARQANIPVLIDPKGTDFERY	186				
Qy	187	RGATLLTPNLSSEFAVGVCKTETEEIVERGMKMLIADYELSALLVTRSQGMSLLQPGKAP	246				
Db	187	RGATLLTPNMSEFAVGVCKDSEEEIEKGLKLEIQIDLSSALLVTRSEKGMTLLRPNQPA	246				
Qy	247	LHMTQAOEYVDVTGADGTVTIGVLAATLAAGNSLEEEACFFANAAGVVGKLGTSVTSPI	306				
Db	247	FHLATEAKEVFDVTGADGTVLSVLATGLADGRSEEEACYLANVAAGVVGKLGTSVTSV	306				
Qy	307	ELENVGRADTGFVMTETEBELKLAVAARKKRGKVVMTNGVFDILHAGHSYSLANARKL	366				
Db	307	ELENAIHGRSNTGFGIMSEAEKRWQAQARGKEIKVMTNGCFDILHPGHYSYLENARKL	366				
Qy	367	GDRLIVAVNSDASKRLKGDSPRPNPLORMIVLGALEADVVMVVSFEEDTTPQRLTAGTLP	426				
Db	367	GDRLIVAVNSDDSVKRLKGETRPINDLQSRMAVLAGLSSVDWLVAFHEDTTPQRLAREVLP	426				
Qy	427	DLLVKGDDYKPEETAGSKVWANGEVVLVLNFEDGCGSTNTTIKKIQDOK	475				
Db	427	DLLVKGDDYKPEDTAGSQEYVWASGDDVKVLNLFNGCGSTSNVISIKIQOLK	475				

RESULT 8	
Q9HUG9	
ID	PRELIMINARY; PRT; 474 AA.
AC	Q9HUG9
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	LPS biosynthesis protein RfaE.
GN	RFAE OR PA4996.
OS	Pseudomonas aeruginosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC	Pseudomonas.
OX	NCBI_TaxID=287;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 15692 / PAO1;
RX	MEDLINE=20437337; PubMed=10984043;
RA	Stover C.K., Pham X.-O.T., Erwin A.L., Miziouchi S.D., Warrenner P.,
RA	Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA	Garday R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

[illegible]

```

RESULT 9
Q48046
ID Q48046 PRELIMINARY; PRT; 342 AA.
AC Q48046;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ADP-heptose synthase.
GN RFAE.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=2019;
RC MEDLINE=96070820; PubMed=7592970;
RA Lee N., Sunshine M.G., Engstrom J.J., Gibson B.W., Apicella M.A.;
RT "Mutation of the trb locus of Haemophilus influenzae nontypeable
RT strain 2019 is associated with modification of lipid A and

```

RT phosphorylation of the lipooligosaccharide.";
RL J. Biol. Chem. 270:27151-27159(1995).
RN [2]
RC STRAIN-2019;
RX MEDLINE=95172727; PubMed=7868252;

RA Lee N.G., Sunshine M.G., Apicella M.A.;
RT "Molecular cloning and characterization of the nontypeable Haemophilus
influenzae 2019 rfaE gene required for lipopolysaccharide
biosynthesis.";
RL Infect. Immun. 63:818-824(1995).
DR EMBL: U17642; AAC43516.1; -;
DR InterPro: IPR002173; PfKB.
DR Pfam: PF00294; pfkb; 1.
SQ SEQUENCE 342 AA; 47538 MW; 4FFB55031E3927C8 CRC64;

Query Match 46.9%; Score 1122; DB 2; Length 342;
Best Local Similarity 67.2%; Pred. No. 5.8e-63;
Matches 217; Conservative 47; Mismatches 59; Indels 0; Gaps 0;

QY 7 EFERAGVMVGVDMVDRYWGPTSRISPEAPVPVVKVNTIERPGGAANVAMNIASLGAN 66
DB 7 EFKQAKVLVGVDMVDRYWGATNRISPERPVPVVOENEERAGGAANVAMNIASLNP 66
QY 67 ARVLGLTGIDDAARALSKSLADNVKCDVSVTPHTITKRLVSRNQQLRLDFEFGFE 126
DB 67 VOLMGLIGQDETGSALSHLEKQIDCNFVALETHPTITKRLILSRHQQLRLDFEEDFN 126
QY 127 GVDPOPLHERINQALSIGALVLSYAKGALASVQOMIQIARKAGVPLVDPKGTDFERY 186
DB 127 NVDCDKDLAKLESARKYKALISLDYKGLKDKQKMIQIARKGNVPLVDPKGTDFERY 186
QY 187 RGATLLPNLSEFPAVVGKCTEEIEVERGMKLIADYELSALLVTRSEOGMSLLOPKCAP 246
DB 187 RGATLLPNSEFPAVVGKCTEEIEEIKGLKLSLDELTAALLVTRSEKMTLLRPQEP 246
QY 247 LHMPAQEYVDVTGAGDTVIGVLAATLAAGNSLEAEAFANAAAGVVGKLTSTVSPI 306
DB 247 YHLPVAKVEFVDVTGAGDTVIVSLATLADGCSFEESCYLANVAAGVVGKLTSTVSTV 306
QY 307 ELENVAVGRADTGFVMTBEELK 329
DB 307 ELENIAIHARPETGFGIMSEALK 329

RESULT 10

Q9A2C5 Q9A2C5 PRELIMINARY; PRT; 483 AA.
AC Q9A2C5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RfaE protein.
GN CC3640.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE006022; AAK25602.1; -;
DR HSSP: Q9TPW2; 1DG.

DR TIGR: CC3640; -;
DR InterPro: IPR004820; Cytidylyltransf.
DR InterPro: IPR004821; Cyt tran_rel.
DR InterPro: IPR002173; pfkb.
DR Pfam: PF01467; Cytidylyltransf; 1.
DR Pfam: PF00294; pfkb; 1.
DR TIGRfams: TIGR00125; cyt_tran_rel; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 483 AA; 49621 MW; 9CDAED6CD784608A CRC64;
Query Match 37.8%; Score 904.5; DB 16; Length 483;
Best Local Similarity 45.1%; Pred. No. 5.1e-49;
Matches 214; Conservative 66; Mismatches 186; Indels 9; Gaps 8;
QY 5 LPE-FERAGVMVGVDMVDRYWGPTSRISPEAPVPVVKVNTIERPGGAANVAMNIASL 63
DB 8 LPFAFAGTKVLVGVDMVDRFYGAVDRISPEAPVIAVEKTEFAMTLCGGAGNARNVAAL 67
QY 64 GANARLVGLTGIDDAARALSKSL-ADNVKCDVSVTPHTITKRLVSRNQQLRLDFE 122
DB 68 GAKAVLGLVCRDDAGAALRCMDAEEAGLEAEVLVDPARRTEKRVYISGSHQMLRVDR 127
QY 123 EGFEGVDPQLHERINQALSSIGALVLSYAKGAL--ASVQOMIQIARKAGVPLVDPK 180
DB 128 DRSPG-DGAALLAAAFETRLASADVVLSDYAKGVLPVAVVRGAIDAAKAAGKPVIVDPKS 186
QY 181 TDERYRGATLLPNLSEFPAVVGKCTEEIEE-EGMKLIADYE-LSALLVTRSEOGMS 238
DB 187 RDEARYDGATLIKPNKEAAEATGIVTSDAASDAGAAIILAMAPGLOALITRGAGMT 246
QY 239 LLOPGKAPLHMPAQEYVDVTGAGDTVIGVLAATLAAGNSLEAEAFANAAAGVVGK 298
DB 247 LAVNQPIHLPTAIEVDFVSGAGDTVAATLALAVAGASLAQAQALNAGLVAKL 306
QY 299 GTSTVSPLELEN-AVRGRADTG-FGVMTBEELKLAFAAARKRGKVVMTNGVDFDLHAGH 356
DB 307 GTDVTVAELTACASSAGGEPGEIKIADREQAQRIVEGWARGKLVGTNGCFDILLPGH 366
QY 357 VSYLANARKLGLRLIVAVNSDASTKRLKGSRPVNPLEORMIVLGALEAVDVMVVFEDT 416
DB 367 VLLSQAKAACDRLIVGLNTDASVSKLGTPRVQKQGRATVLAASSVDLVLFDEDT 426
QY 417 PORLIAGILPDLLVKGDDYKPEETAGSKVWANGVEVLVNFEDGCSSTNIKKI 471
DB 427 PLELIKAFRFDVLVKGADYTVETVVGSDVVLGYGKVVLAELKQGSQSTNLIARM 481

RESULT 11

Q9PNE5 Q9PNE5 PRELIMINARY; PRT; 461 AA.
AC Q9PNE5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative ADP-heptose synthase.
GN WAAE OR CJ1150C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).

```
DR EMBL; AL139077; CAB73404.1; -.
DR HSP; P27623; 1CO2.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR Pfam; PF00294; pfkb; 1.
DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
KW Complete proteome.
SQ SEQUENCE 461 AA; 51237 MW; 490D9FDB4EB58322 CRC64;

Query Match 36.1%; Score 863; DB 16; Length 461;
Best Local Similarity 43.1%; Pred. No. 2e-46;
Matches 203; Conservative 90; Mismatches 158; Indels 20; Gaps 8;

QY 9 ERAGVAVGVMDLDRWYGTPTSRIPEAPVPVVKVNTIERPGCAANVANIASLGANAR 68
DB 8 QKPLIIGDFVMDNTWDCSRISPEAPVLIATKEDKRLGCAANVANIASLGADV 67
QY 69 LVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKRLVLSRNOQLIRLDPFEGV 128
DB 68 ALGVGDDESGKFLQE-----NLKGEFLIOKGRKTFKKNRIMAHNQVLRLEDEISEIL 122
QY 129 DPOPLHERINQALUSSGALVLSYAKGALAS--VOQMQLARKAGVPLIDPKGTDFERY 186
DB 123 LENELIALFDEKIDPKAVVLSYAKGVLTPKVKAKVIEKAKVNLINPVLVDPKGSDFNKY 182
QY 187 RGATLTPNLSEFAVVGCKTEEEIVERGM-KLIADYELSLVTRSEOGMSLLQPGKA 245
DB 183 SGATLTPNKKALEKLENLEGENLEKIKLKEDEFSLRYSITLSEAGIALFDEGLK 242
QY 246 PLHMPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAGVVGKLGSTVSP 305
DB 243 --TAPAKALEYDVTGAGDSVIAVLAFLANEIEIFKACELANEAAVVGKIGSVSVSF 300
QY 306 IELENVGRADTGFQVMTTEELKLAFAAARKGKGVMTNGVDFDILHAGVSVYLANARK 365
DB 301 DEIKSF--KRVDPFKIKSKEEL---LVLLKQNNKIVETNGCFDVFHFGHIKYLDAKR 355
QY 366 LGRLLVAVNSDASTKRLGDSRPVPLEQRMIVLGALEAVDWMVSEEDTPQRLTAGIL 425
DB 356 LGRLLVGLNSDASVRLKGESRPVNSEFORACMLAAFYVDFVDFDTPLELISFLK 415
QY 426 PLLLVKGGDYKPEIAGSKREVWANGGEVLVNFEDGCGSTTNIKKIQODKK 476
DB 416 PDILVKGADYKDLVVGADIV-----SRVELIDFEEGFSSTKIIEKI-KOKK 461

RESULT 12
Q98154 PRELIMINARY; PRT; 496 AA.
AC Q98154;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE ADP-heptose synthase.
GN MLL2562.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFE303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
```

```
DR EMBL; AP003000; BAB49662.1; -.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR Pfam; PF00294; pfkb; 1.
DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
KW Complete proteome.
SQ SEQUENCE 496 AA; 51449 MW; 49DB1E271653AA76 CRC64;

Query Match 36.1%; Score 863; DB 16; Length 496;
Best Local Similarity 43.3%; Pred. No. 2.2e-46;
Matches 209; Conservative 69; Mismatches 183; Indels 22; Gaps 8;

QY 5 LPFERAGVAVGVMDLDRWYGTPTSRIPEAPVPVVKVNTIERPGCAANVANIASLG 64
DB 16 IAREFGVTVLVVGDGLDILDRFVNGVIERISPEAPIPVLHGRGETSAMGGAGNVANISLG 75
QY 65 ANARLVGLTGIDDAARALSKSLADVNVKDFVSVPT-HPTITKRLVLSRNOQLIRLDFEE 123
DB 76 ARAIPYSVIGTDTAGDSLVRMLAELGAETAGLSQQRGRMTSSKSRESALNQVLRDEEE 135
QY 124 GFEGVDPQPLHE-----RINQALSSGALVLSYAKGALAS--VOQMQLARKAGVP 173
DB 136 -----IKPLDETERAGLIRHFRALAGAEIVILSDYKGILLDGVAAELIACREAGKP 189
QY 174 VLIDPKGTDFERYRGATLTPNLSEFAVVGCK--TEEEIVERGMKLIADYELSLVTR 232
DB 190 VLVDPKGRDYARVAGATATIPNKELGAEVGHAFVADDEIVAAARELISHAGDFVYATR 249
QY 233 SEGMSLLQPGKAPLHMPTQAOEYDVTGAGDTVIGVLAATLAAGNSLEAEACFFANAAAG 292
DB 250 SEGKMSVVGPDGA-RHIATQAREVDFVSGADTVIATFALALASGADPVAASASTANAAGG 308
QY 293 VYVGKLGSTVSPVIELENV-RGRADTGF--GVMTTEELKLAFAAARKGKGVMTNGVNF 349
DB 309 VYVGKLGSTVSPVIELENV-RGRADTGF--GVMTTEELKLAFAAARKGKGVMTNGVNF 368
QY 350 DILHAGVSVYLANARKLGRLLVAVNSDASTKRLGDSRPVPLEQRMIVLGALEAVDWMV 409
DB 369 DILHAGVSVYLANARKLGRLLVAVNSDASTKRLGDSRPVPLEQRMIVLGALEAVDWMV 428
QY 410 VSEEDTPQRLTAGILPDILVKGADYKPEIAGSKREVWANGGEVLVNFEDGCGSTTNIK 469
DB 429 VSEEDTPQRLTAGILPDILVKGADYKPEIAGSKREVWANGGEVLVNFEDGCGSTTNIK 488
QY 470 KIQ 472
DB 489 KLR 491

RESULT 13
Q9X518 PRELIMINARY; PRT; 313 AA.
ID Q9X518;
AC Q9X518;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Putative ADP-heptose synthetase (Fragment).
GN RFAE.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RA Kahler C.M., Stephens D.S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF125564; AAD32179.1; -.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfkb; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
DR NON_TER 1
```


Db 14 FAQAKVLVVGDMYDRYWFVGDVSRISPEAPVPVAKIGRIDORAGGAANVARNIASLGGKV 73
Qy 68 RLVLGTGIDDAARALSKSLADVNVKDFVSPTHTITKLRVLSRNOQLRLDFEEGEG 127
Db 74 GLSVTGNDERAADALDALMVQDGVASYLMRDQKIATTVKLRVARNQQLRLDFEEHPNR 133
Qy 128 VDPQPLHERINOALSSIGALVLSDYAKGALASVQOQMIQARKAGVPVLIDPKGTDFFERYR 187
Db 134 EVLEQIKRKYREILPEYDAIIFSDYKGGSLSHISDMIDWAKHEGKTVLIDPKGDDYKYA 193
Qy 188 GATLLTPNLSEFEAVVGCKTEEEIIVERGMKLIADYELSALLVTRSEOGMSLLQPGKAPL 247
Db 194 GATLLTPNRAELKEVVGSKNENDLTERAQNLRRLDLTALLTRSEEGMTLFSEGE-PI 252
Qy 248 HMPTQAEVYDVTGAGDVIIGVLAATLAAGNSLEEACFFANAAAGVVYVKLGTSVSPTE 307
Db 253 YOPTRAQEVYDVSAGDVIAGMGLGLAAGCTMPEAMYLANTAAGVVVAKLGTAVCSFAE 312
Qy 308 LENA VRGRA 316
Db 313 LTKALSGQS 321

Search completed: November 24, 2002, 23:18:30
Job time : 98 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2002, 23:14:19 ; Search time 37 Seconds
(without alignments)
379.317 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERAGVWVGDVY.....FEDGCSYTIKIKIQDCKG 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	165.5	6.9	315	4	US-09-134-001C-4521
2	150	6.3	137	4	US-09-134-001C-5560
3	137.5	5.7	312	4	US-09-134-001C-5563
4	137.5	5.7	347	3	US-08-826-611-2
5	134.5	5.6	3724	2	US-08-804-227C-10
6	134.5	5.6	3724	2	US-08-804-198-4
7	130	5.4	859	4	US-09-199-637A-281
8	117.5	4.9	362	2	US-08-479-614-8
9	116.5	4.9	361	2	US-08-479-614-2
10	116	4.8	328	3	US-08-826-611-6
11	115.5	4.8	366	4	US-08-955-957A-3
12	112.5	4.7	303	2	US-08-961-539-2
13	112.5	4.7	303	4	US-09-185-826-2
14	111.5	4.7	345	2	US-08-479-614-5
15	110	4.6	321	4	US-09-134-001C-3900
16	110	4.6	3739	3	US-09-320-878-2
17	110	4.6	3739	4	US-09-105-537-33
18	110	4.6	11877	4	US-09-105-537-6
19	107.5	4.5	507	4	US-09-091-097-34
20	106	4.4	547	4	US-08-461-722-2
21	106	4.4	547	4	US-08-336-251-2
22	106	4.4	547	5	PCT-US94-06362-2
23	104	4.3	4150	4	US-09-428-517-2
24	103.5	4.3	1056	2	US-08-627-873-7
25	102.5	4.3	552	1	US-08-116-098-2
26	102.5	4.3	552	4	US-08-687-590-32
27	102	4.3	1861	2	US-08-790-912-4

28	101.5	4.2	530	3	US-09-222-817-12	Sequence 12, Appl
29	101.5	4.2	530	4	US-09-222-786-12	Sequence 12, Appl
30	101	4.2	6095	4	US-09-144-085-2	Sequence 2, Appl
31	100	4.2	1864	2	US-08-804-227C-3	Sequence 3, Appl
32	100	4.2	5087	4	US-09-144-085-1	Sequence 1, Appl
33	100	4.2	7257	3	US-09-335-409-5	Sequence 5, Appl
34	100	4.2	7257	4	US-09-568-102-5	Sequence 5, Appl
35	100	4.2	7257	4	US-09-567-969-5	Sequence 5, Appl
36	100	4.2	7257	4	US-09-568-480-5	Sequence 5, Appl
37	100	4.2	7257	4	US-09-568-486-5	Sequence 5, Appl
38	100	4.2	7257	4	US-09-568-472-5	Sequence 5, Appl
39	100	4.2	7257	4	US-09-567-899-5	Sequence 5, Appl
40	99.5	4.2	1313	4	US-09-071-035-450	Sequence 450, App
41	99.5	4.2	1313	4	US-09-071-035-454	Sequence 454, App
42	99.5	4.2	1891	2	US-08-804-227C-12	Sequence 12, Appl
43	99.5	4.2	1891	2	US-08-804-198-6	Sequence 6, Appl
44	99.5	4.2	3567	2	US-07-642-734C-4	Sequence 4, Appl
45	99.5	4.2	3567	3	US-08-439-009A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-134-001C-4521
; Sequence 4521, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4521
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4521

Query Match	6.9%	Score 165.5;	DB 4;	Length 315;
Best Local Similarity	22.6%	Pred. No. 3.1e-08;		
Matches	77;	Conservative	56;	Mismatches 142;
				Indels 65;
				Gaps 13;
Qy	2	KVTLPEFERAGVWVGDVYWGPTSRISPEAPVPVVKVNTIEERPGG--AANVAMN	59	
Db	3	KVGEVEMKNKVIIGSTNVDFL--NVKRPK-PGETLHINQAKREFGGKGANQAI	58	
Qy	60	IASIGANARLVGLGIDDAARALSADVNWKDFVSPVTHPTITKLVLSRNOQLRL	119	
Db	59	ASRLAADTFTISKVKDGNANFILE-----DFKKGAIH-----TQYILTS	98	
Qy	120	DFEE--GFEQVDPQP-----LHERINOALS-----SIGALVLSDYAKGAL--ASVQ	161	
Db	99	ESEETGQAFITVDEAGONTILVGGANNTLSATDVMSVDFAIGADFVVAOLEVPFAIE	158	
Qy	162	QMIQLAKAGVPVLIDP-----KGTDFERYRGATLLTPNLSEFAVVGCKTEE	210	
Db	159	QAFKIARKQNTTVNLNPAIPLPKSLLELTD-----IIIPNEIEALLTGISINNE	210	
Qy	211	EIVERGKMLIADYELSALLVTRSGMSLLQPGKAPLHMPQAOEVDVTCAGDTVIGVL	270	
Db	211	SDMKETATYFLDLGTSVAVLITLGEQGYCAYOEQVKM-IPACNVKAIDTTAAGDTFIGAF	269	
Qy	271	AATLAAG-NSLEECFFANAAAGVVVVKLGITSTVSTPIELE	309	
Db	270	LSELNKDLSNLESIRLANQASSLTVQRKGAQASIPTRKE	309	

```
RESULT 2
US-09-134-001C-5560
; Sequence 5560, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5560
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5560

Query Match      6.3%; Score 150; DB 4; Length 137;
Best Local Similarity 33.8%; Pred. No. 2.8e-07;
Matches 47; Conservative 29; Mismatches 49; Indels 14; Gaps 6;

QY 341 KVMYNGVFDILHAGHVSFLANARKLGDRLIVAVNSDASTKELKGDSPVNPLEQRMIVL 400
DB 7 KRVIYGYDILLHYGHIELLRAREMGDYLVALSTD-EFNOIK-NKKSYYDYEQKMWL 64

QY 401 GALEAVDWVVSFEETPQRL--IAGTLPDLLVKGDDYKPEETAGSKYEWANGGEVLVNF 458
DB 65 ESIRYVDLVIP-EEGWQKEKDVRFDVDFVGMHDWE-----GEFDLKKDCEVILNR 118

QY 459 EDGCSTTNIKKIQDKKG 477
DB 119 TEGISTT-----KIKOELYG 133

RESULT 3
US-09-134-001C-5583
; Sequence 5583, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5583
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5583

Query Match      5.7%; Score 137.5; DB 4; Length 312;
Best Local Similarity 21.6%; Pred. No. 2e-05;
Matches 58; Conservative 58; Mismatches 111; Indels 41; Gaps 12;

QY 51 GGAANVAMNIASLGANARLVGTGIDDAARALSKSLADVNVKDFVSVPTHTTKLRVL 110
DB 43 GKGINVSRLVLTLDVSTALGFSG-GFGDFTIAQTLEDNSIQSDFFQVDED---TRINKV 98

QY 111 SRNQLIRLDFEEGEGVDPQPLHERINQALSSI-----GALVLSDYAKGALAS--VQO 162
DB 99 LKSGQ-----ETEINAPGPKVTHAQFEQLLSQIRRTTNDIDIVIVAGSVNPSPDAYAQ 152
```

```
QY 163 MIQLARKAGVPVLIDPKGTDFER---YRGATLLTLPNLSEFEAVVG-KCKTEEEIVERGMK 218
DB 153 IAQITEKTGAQLVDAEKDLVETVLPYR-PLFIKPKDLEVMFNTTVKSDSDVIKYGKE 211
QY 219 LIADYELSALLVTRSEQGMSSLLQPKAPLHMTQA---QEVYDVTGAGDVTGIVLAATL 274
DB 212 ILKKGAAQSVIISLGDGAIYVDQ-----HQSIKAVNPQGHVYVNVGSDSTVAGMVAGL 265
QY 275 AAGNSLEACFFANAAAGVVVGKLGST 302
DB 266 SMGLNIDEA--FQAVAS-----GTAT 285

RESULT 4
US-08-826-611-2
; Sequence 2, Application US/08826611
; Patent No. 6031154
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B.
; APPLICANT: Kanayama, Yoshinori
; TITLE OF INVENTION: Fructokinase Genes and Their Use in
; TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,611
; FILING DATE: 05-APR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-07740005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-611-2

Query Match      5.7%; Score 137.5; DB 3; Length 347;
Best Local Similarity 24.4%; Pred. No. 2.4e-05;
Matches 85; Conservative 48; Mismatches 131; Indels 85; Gaps 18;

QY 13 VMVGDVMDRYVGYGTSRISPEAPVPVVKVNTIETPCCA-ANVAMNIASLGANARLVG 71
DB 28 VYCFGEMLID---FIPT-----VAGVSLAEPAPFAKPGCAPANVAVCSIKLGSSAFIG 79
QY 72 LTGIDDAARALSKSLADVNVKDFVSVPTHTTKLR-----VLSRN---QOL 116
DB 80 KVGDDFEGRLADILKQNVNDNSGHRF-DHDARTALAFITLTAEGEREFVFFRNPADML 138
QY 117 IR---ID-----FEEGEGVDPQPLHERINQALSSIGALVLSDYAK--GALASVQO 162
DB 139 LRESELDVLDLIKATIFHYGSIISLIDPCRS-----HLAAMDIAKSGSILSDP 189
QY 163 MTQL-----ARKAGVPVLIDPKGTDFRYGATLLTLPNLSEFEAVVG-KCKTEEEIVE 214
DB 163 MTQL-----ARKAGVPVLIDPKGTDFRYGATLLTLPNLSEFEAVVG-KCKTEEEIVE 214
```

Db 190 NLRPLWPSDAARSGI-----MSVNLADIKIKEDEISFLTGADPNDEVV- 238
QY 215 RCMKLIADYELSAIVTSEQMSLLQ---PKADLHPTQAEVYDVTGAGDVTIGVLA 271
Db 239 --LKRFLHNPULKLLVTTGSGACRYTKEFKGR-----VNSIKVKAVDTTGAGDAFTGGVL 292
QY 272 ATLAAGNS-----LEACFFANAAAGVYVKGKLTSTVSPTELENAVR 313
Db 293 KCLASDASLYQDEKRLREAIFFANVCAALTV--TCRGGIPSLPTQDAVR 339

RESULT 5
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-10

Query Match 5.6%; Score 134.5; DB 2; Length 3724;
Best Local Similarity 24.0%; Pred. No. 0.0026;
Matches 112; Conservative 48; Mismatches 196; Indels 111; Gaps 20;

QY 10 RAGVWVGDMVLDRYWYGTSTRISPEAPVPVVKVNTIERPGCGAANVANNTIASGANARL 69
Db 1178 RGTVLIAAGL-----TTVPGRVLSLLE--DGADRVVLAGDPAPAAAA 1219
QY 70 VCLTGI-----DDAARALSKSLADVNVKDFVSVPTHTITKLRVLSRNQQLRLDPE 122
Db 1220 AGLTGVSLVPVRCDDTDRALAAALLDEHAPTAVVHAPPLVPLAPURETAPGDIAAALAK 1279
QY 123 EGFEG--VDPOPLHERINQALSSIGALVLSYAKG-----ALASVQOQMIQLAR 168
Db 1280 TTAAGHLVDLAP-----AAGLDALVLFSSVSGWGGAAGGYAASAHLDALAERAR 1331
QY 169 KAGVPVL-----IDPKGDFERYRGATLLTNLSEFAVVGKCKTEEIVEIRG 216
Db 1332 AAGVPAFVAVSWPAGGTPADGAEEFLSRRLGLAPLDPD---QAV-----RTLRRMLERG 1383
QY 217 MK--LIADYELSAIVTSEQMSLLQPG---KAPLHPTQAEVYDVTGAGDVTIGVL 270

Db 1384 SACGAVADVENS-----RFAASYTWVRPAVLPDDLPDQVRLRAAEAPSTGSTSELVR 1438
QY 271 AATLAAGNSLEPACF-FANAAAGVVVGKLGSTVSPIELENAVRGADTGFVGMTEEEL- 328
Db 1439 ELTAQSGHKRHATLRLVRHAHAAAVLGSSGDVAVS-----SARAFRLGDFSLTALELR 1492
QY 329 -KLAVAAAHKRGEKVVMTNGVFDILHAGHVSYLANARKLGDRLIIVAVNSDASTKRLKGDS 387
Db 1493 DRLSTSTGLK-----LPTSLVFD-----HSSPAALAHRLHGEELL-----GRNDTADRAGPD 1538
QY 388 RPNVPLEQRMIVLGALEAVDWWVYFEEPTPQRLIAGILPDLLVKGGD 434
Db 1539 TPVR-TDEFIATIGMACRLPGGVQSPED-----LWDLITGGTD 1575

RESULT 6
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-198-4

Query Match 5.6%; Score 134.5; DB 2; Length 3724;
Best Local Similarity 24.0%; Pred. No. 0.0026;
Matches 112; Conservative 48; Mismatches 196; Indels 111; Gaps 20;

QY 10 RAGVWVGDMVLDRYWYGTSTRISPEAPVPVVKVNTIERPGCGAANVANNTIASGANARL 69
Db 1178 RGTVLIAAGL-----TTVPGRVLSLLE--DGADRVVLAGDPAPAAAA 1219
QY 70 VCLTGI-----DDAARALSKSLADVNVKDFVSVPTHTITKLRVLSRNQQLRLDPE 122
Db 1220 AGLTGVSLVPVRCDDTDRALAAALLDEHAPTAVVHAPPLVPLAPURETAPGDIAAALAK 1279
QY 123 EGFEG--VDPOPLHERINQALSSIGALVLSYAKG-----ALASVQOQMIQLAR 168
Db 1280 TTAAGHLVDLAP-----AAGLDALVLFSSVSGWGGAAGGYAASAHLDALAERAR 1331

```

Query Match      5.4%; Score 130; DB 4; Length 859;
Best Local Similarity 23.8%; Pred. No. 0.00062;
Matches 123; Conservative 75; Mismatches 178; Indels 140; Gaps
29;

Qy 39 PVVKVNTIEPRGCA-----ANVAMNTASL--GANARLVGLTGI-----DDAARA 81
|| : : || : || : || : || : || : || : || : || : || : || : ||
Db 36 PVHLSALLEQGGGSIKPLMQGVFDIAALRSGLNGELDALPKIQSPTGDYNLSQDLARL 95
|| : : || : || : || : || : || : || : || : || : || : || : ||
Qy 82 LSKS-----LADNVNCKDFSVTPHTPTIKRLVLSRNOQLIRLDFE-----EGFEGVD 129
|| : : || : || : || : || : || : || : || : || : || : || : ||
Db 96 LNQADRLAQQKGDQFISSLVLLAAMDENTRKLGLLGQGVSRKALENAVANLRGGEAVN 155
|| : : || : || : || : || : || : || : || : || : || : || : ||
Qy 130 PQPLHERINQALSSIGALVLSDYAKGAL-----ASVQOQIQL--ARKAGVPVLIDPKG 180
|| : : || : || : || : || : || : || : || : || : || : || : ||
Db 156 -DPNVEESRQALDKYTVDMTKRAEEGKLDPVIGRDELRRTIQVLQRTKKNPVLI GEPG 214
|| : : || : || : || : || : || : || : || : || : || : || : ||
Qy 181 TDFERYRGATLLTPNLS-----EFAVVGCKCTEETEEIVRGMKLI 220
|| : : || : || : || : || : || : || : || : || : || : || : ||
Db 215 V-----GKTAIVEGLAQRIINGEVPDGLKDKRLLALDMLAGAKPRGEFEER----- 263
|| : : || : || : || : || : || : || : || : || : || : || : ||
Qy 221 ADYELSALLVTRSQGNSLLQPGKRAPLHMPQAOEVDYDVTGAGDTVIGVLAATLAAGNSL 280
|| : : || : || : || : || : || : || : || : || : || : || : ||
Db 264 ----LKAVL----NELGK---QEGRVILFI---DELHTTMVAGRA-----EGAMDAGNML 304
|| : : || : || : || : || : || : || : || : || : || : || : ||

```

```

Query Match      4.9%; Score 117.5; DB 2; Length 362;
Best Local Similarity 19.0%; Pred. No. 0.0026;
Matches 70; Conservative 55; Mismatches 145; Indels 99; Gaps

Qy 1 MKVTLPEERAGVM-----VGDVMDLRVWYGPTSRISPEAP-----VPPVKV 43
      ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 13 LKVEAPQALRENILFGMGNPLLDISAYVDKDFDLKYSKPNQOILAEADKHKEFLDELVRK 72
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 44 NTIEERCGGAANVAAMTSLGAN-----ARLVGLTGIDDAARALSRLADVNVYKDFVSV 98
      ||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 73 FKVEYHAGGSQTSIKVQAWMIQOPHKAATFFCGICIGIDFEGEILKKRAAEAHVDHAYEQ 132

```

```

QY   99 PTHPTT-KLRVLSRNQOLI-----RLDFEES-----PEGVD 120
      ||| | :|| | || |
Db   133 NEOPTGTCAACITGDRNSLIANLAANCYKKEKHLDLEKNMMLVEKARVCYTAGFFLTYS 192
      |:|::---HERINQAAL--SSIGALVLSDYAKGALASVOQMIOQL-----ARKAG 171
QY   130 POPL-----
      |:|:::~::~|||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   193 PESVLKVAAHASENNRIFTLNLSAPFISOFYKESLMKMVPYDYLFGNETEAATFAREQG 252
      :|:~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:
QY   172 VPVLIDPKGTFDYRYRGATLLTNL-SEFEAVVGCKCTBEEIIVERGMKLIAOYELSALLV 230
      :~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:
Db   253 F-----ETDKIEAIKKTAOLPKMSKRQRIVIFTGRDDTI-----MATSEVTAFAY 301
      :~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:
QY   231 TRSEQGMSLLQPCKAPLPHMPTQAQEYDVDTGADTVGVGLAATLAAGNSLEEACFPANAA 290
      :~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:
Db   302 LDQDQ-----KEIIDTNGAGDAEVGGFLSQLVSVDKPLTECIRAGHYA 343
      :~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:
QY   291 AGVVVKLG 299
      |:::|
Db   344 ASIIIRRTG 352

RESULT 9
US-08-479-614-2
; Sequence 2, Application US/08479614
; Patent No. 5861294
; GENERAL INFORMATION:
; APPLICANT: Cowart, Marlon Daniel, Halbert, Donald N.,
; APPLICANT: Kerwin, Jr., James F., McNally, Teresa
; TITLE OF INVENTION: Adenosine Kinase Polypeptides
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: D-377 AP6D, 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System 7.1
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,614
; FILING DATE: June 7, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas D. Brainard
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5749.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 937-4884
; TELEFAX: (708) 938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-479-614-2
```

```

Qy 44 NTTEERPGGAANYAMNTASLGN-----ARLVGLTGTDIDDAARALSKSLADVNVVKDFSV 98
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 72 FKVEYHAGGQTSMKVAOWMIQEPHRAATFFCGICIGDKFGEILKSKAADAHVDHYEQ 131
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Qy 99 PTHPTIT-KURLVLSRNOOLI-----RLDPEEG-----PEGVD 129
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 132 NEQPTGTCACTGGNRSLSVANLAANCYKKEHLDELNNWMLVEKARYVYTAGFFLTVS 191
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Qy 130 POPLHERINQALS-----SIGALVLSDYAKGALASVOQMIOI-----ARKAG 171
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 192 PESVLKVARYAAENNRFTTLNSAPFTSQPFQKEALMEVPYVDILFGNETEAATFAREQ 251
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Qy 172 VPULIDPKGDIFERYRGATLLTP--NLSEEEAVVGCKTETEEEIVERGMKLIADYELSALL 229
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 252 F-----ETKDIEIAKRTQALPKVNSKRQRTVIFTQGRDDTIATG-----NDVTAF 299
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Qy 230 VTRSEQMSLLQPCAKPLHMPTQAQEYDVDTGAGDTVIGVLAATLAAGNSLEACFFANA 289
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 300 VLQNQ-----EIVDTNGADGAFVGGFLSQLVSNKPLTECIRAGHY 341
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Qy 290 AAGVVVGKLG 299
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :
Db 342 AASVIIRRTG 351
   :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :

RESULT 10
US-08-826-611-6
; Sequence 6, Application US/08826611
; Patent No. 6031154
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B.
; APPLICANT: Kanayama, Yoshinori
; TITLE OF INVENTION: Fructokinase Genes and Their Use in
; TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,611
; FILING DATE: 05-APR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0774000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-611-6

```


[illegible]

```

RESULT 14
US-08-479-614-5
; Sequence 5, Application US/08479614
; Patent No. 5861294
; GENERAL INFORMATION:
; APPLICANT: Cowart, Marlon Daniel, Halbert, Donald N.,
; APPLICANT: Kerwin, Jr., James F., McNally, Teresa
; TITLE OF INVENTION: Adenosine Kinase Polypeptides
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: D-377 AP6D, 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System 7.1
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,614
; FILING DATE: June 7, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas D. Brainard
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5749.05.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 937-4884
; TELEFAX: (708) 938-2623
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-479-614-5

```

Query Match	4.7%	Score 111.5	DB 2	Length 345
Best Local Similarity	19.2%	Pred. No. 0.0096		
Matches	66	Conservative 51	Mismatches 139	Indels 87
Gaps	12			
Qy 15	VVGDVMLDRVYWGPTSRISPEAP-----VPVVKYNTTEERPGGAANVAMNTASLGAN----	66		
Db 22	VVDKDFDLKYSKPNDOQILAEKHKELFDELVLKFKVEYHAGGSTGTSIKVAQNMIOOPH 81			
Qy 67	--ARLVGLTGIDDAARALSLSADVNVKDFVSVPTPTIT-KLVLSRNQOLI----- 117			
Db 82	KAATFFCGIGDKFGKILKRAAEAHVDAHYEONEQPTGTCACTIGDNRSLTANLAAA 141			
Qy 118	-----RLDFEFG-----FEGVDPPQL-----HERINQAL-----SSIGALV 148			


```

Db 142 NCYKKEKHLDEKKNMLVEKARVCYIAGFFLTVSPESVLKVAHASENNRIFTLNSAPP 201
QY 149 LSDYAKGALASVQOMIQL-----ARKAGVPVLIDPKGTDFFERYGATLLTPNL- 196
Db 202 ISOFYKESLMKVMYPVDILFGNETEAATFARQGF-----ETKDIKEIAKKTQALPKMN 255
QY 197 SEFEAVVGKCTEEIEVERGMKLIADYELSALLVTRSEQGSLLQPGKAPLHMPMQAOEV 256
Db 256 SKQRIVIFTQRDDTI-----MATESEVTAFAVLDDQ-----KEI 292
QY 257 YDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVVGKLG 299
Db 293 IDTNGAGDAFVGGLSQLVSKPLTECIRAGHYAASIIIRRTG 335

RESULT 15
US-09-134-001C-3900
; Sequence 3900, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3900
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3900

Query Match 4.6%; Score 110; DB 4; Length 321;
Best Local Similarity 17.6%; Pred. No. 0.012;
Matches 70; Conservative 53; Mismatches 109; Indels 166; Gaps 16;

QY 13 VMVVGDMLDRYWYGTSRISPEAPVPV---KVNTIEE---RPGGA-ANVAMNIASLGA 65
Db 6 LFAIGEALIDE-----IPNVTHSKLKDVEQFSRQVGAPCNVAATVSKLGG 51
QY 66 NARVLGLTGIDDAARALSKLADVNVKDFVSVPHPTTKLVLRSNQQILRLDFEEGF 125
Db 52 KSEMITQLGND-----AFGDIIVE-----TIEQL----- 75
QY 126 EGVDPQPLHERINOALSSIGALVLSDYAKGALASVQOMIQLARKAGVPVLIDPKGTD--- 182
Db 76 -GVGTQYI-KRTNKANTALAFVSLQD-----DCQDFSFYRKFPSTDMLYOPENIDDIQ 126
QY 183 -----FERYGATLLTPNL----- 196
Db 127 IFQDDILHFCSDVLDIESDMKNAHEKMFESVGGTIVDPNVRLPLWEDKLECORTINA 186
QY 197 -----SEFEAVVGKCTEEIEVE---RGMKLIADYELSALLVTRSEQGSLLQ 241
Db 187 FIPKAHIVKISDEELLFITGK-KNEDEATQSLFRG-----QVNVVIYTOGAQGATIYT 238
QY 242 PGKAPLHMPMQAOEVYDVTGAGDTVIGVLAATLAAGNSLE-----EACFFANAAA 291
Db 239 KDDYRIHHEGYOVQADITTTGAGDAFIGAIYCIILSRHSECKDLFKFKGKOILAFSNRVA 298
QY 292 GVVVGKLGTSVTSPIELENAVRGRADTGFGVMTTEEELK 329
Db 299 ALTTTKHGAIESLP-----TKEDIK 318
```

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2002, 23:15:54 ; Search time 31 Seconds

(without alignments)

240.985 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLPEFERAGVGVGVDM.....FEDGCSSTNIKKIQDDKKG 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2393	100.0	477	10	US-09-741-669-469
2	2393	100.0	477	10	US-09-912-020-325
3	2393	100.0	477	10	US-09-815-242-10301
4	2265	94.7	477	10	US-09-815-242-13759
5	1350.5	56.4	474	10	US-09-815-242-12079
6	197	8.2	308	10	US-09-815-242-11854
7	191.5	8.0	305	10	US-09-815-242-10892
8	178	7.4	306	10	US-09-815-242-11035
9	169	7.1	309	10	US-09-741-669-479
10	169	7.1	309	10	US-09-815-242-10406
11	167	7.0	310	10	US-09-815-242-13976
12	165	6.9	309	10	US-09-815-242-13883
13	158	6.6	309	10	US-09-815-242-10180
14	145	6.1	132	10	US-09-815-242-5399
15	145	6.1	132	10	US-09-815-242-12303
16	145	6.1	132	10	US-09-815-242-12875
17	145	6.1	132	10	US-09-815-242-13151
18	139.5	5.8	304	10	US-09-815-242-13017
19	139.5	5.8	304	10	US-09-815-242-13153

20	130	5.4	854	10	US-09-815-242-5192	Sequence 5192, Ap
21	129.5	5.4	313	10	US-09-815-242-11024	Sequence 11024, A
22	115.5	4.8	134	10	US-09-815-242-10513	Sequence 10513, A
23	115	4.8	1616	9	US-09-712-363-262	Sequence 262, App
24	114.5	4.8	740	10	US-09-815-242-10376	Sequence 10376, A
25	111.5	4.7	303	10	US-09-815-242-13340	Sequence 13340, A
26	110.5	4.6	856	10	US-09-815-242-11134	Sequence 11134, A
27	110	4.6	3739	9	US-09-860-846-33	Sequence 33, Appl
28	110	4.6	3739	10	US-09-861-289-33	Sequence 33, Appl
29	110	4.6	11877	9	US-09-860-846-6	Sequence 6, Appl
30	110	4.6	11877	10	US-09-861-289-6	Sequence 6, Appl
31	109.5	4.6	866	12	US-10-007-693-113	Sequence 113, App
32	109	4.6	396	9	US-09-971-536-64	Sequence 64, Appl
33	108	4.5	858	10	US-09-815-242-11396	Sequence 11396, A
34	106	4.4	627	9	US-10-047-260-36	Sequence 36, Appl
35	106	4.4	844	10	US-09-815-242-11497	Sequence 11497, A
36	106	4.4	1164	10	US-09-984-711-6	Sequence 6, Appl
37	105.5	4.4	308	10	US-09-815-242-10775	Sequence 10775, A
38	105	4.4	302	10	US-09-815-242-4955	Sequence 4955, Ap
39	104.5	4.4	1165	9	US-10-075-460-6	Sequence 6, Appl
40	104.5	4.4	1165	10	US-09-887-052-2	Sequence 2, Appl
41	104.5	4.4	1165	10	US-09-887-052-4	Sequence 4, Appl
42	104.5	4.4	1165	10	US-09-887-052-6	Sequence 6, Appl
43	103	4.3	618	10	US-09-934-901-18	Sequence 18, Appl
44	103	4.3	618	10	US-09-934-868-8	Sequence 8, Appl
45	103	4.3	930	10	US-09-815-242-11133	Sequence 11133, A

ALIGNMENTS

RESULT 1

US-09-741-669-469

; Sequence 469, Application US/09741669

; Patent No. US20020022718A1

; GENERAL INFORMATION:

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; TITLE OF INVENTION: Genes identified as required for

; TITLE OF INVENTION: proliferation of E. coli

; FILE REFERENCE: ELITRA.009A

; CURRENT APPLICATION NUMBER: US/09/741.669

; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: US 60/173005

; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 481

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 469

; LENGTH: 477

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-741-669-469

Query Match 100.0%; Score 2393; DB 10; Length 477;
 Best Local Similarity 100.0%; Pred. No. 3.7e-191;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTLPEFERAGVGVGVDMVLDYWGYSRISRPAPVPVVKVNTIEERPGGAANVANI 60
 |||||
 Db 1 MKVTLPEFERAGVGVGVDMVLDYWGYSRISRPAPVPVVKVNTIEERPGGAANVANI 60

Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSPVTHPTITKRLVLSRNOQLRLD 120
 |||||
 Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSPVTHPTITKRLVLSRNOQLRLD 120

Qy 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOMIOLARKAGVPVLIDPKG 180
 |||||
 Db 121 FEEGEGVDPQPLHERINQALSSIGALVLSYAKALASVQOMIOLARKAGVPVLIDPKG 180

Qy 181 TDFERYRGATLTPNLSEFEAVGVCKTEEEIIVRGMKLIADYELSAIIVTRSEQMSLL 240
 |||||
 Db 181 TDFERYRGATLTPNLSEFEAVGVCKTEEEIIVRGMKLIADYELSAIIVTRSEQMSLL 240

QY 241 QPKAPLHMPDQAEVYDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVGKLG 300
Db 241 QPKAPLHMPDQAEVYDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVGKLG 300
QY 301 STVSPTELENAVGRADTGFGVMTTEELKLAAGNSLEEACFFANAAAGVVGKLG 360
Db 301 STVSPTELENAVGRADTGFGVMTTEELKLAAGNSLEEACFFANAAAGVVGKLG 360
QY 361 ANARKGLDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALAEVDVWVVSFEEDTPQRL 420
Db 361 ANARKGLDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALAEVDVWVVSFEEDTPQRL 420
QY 421 IAGILPDLLVKGDDYKPEETAGSKEYWANGGEVLVNFEDGCSSTTNIKKIQDCKG 477
Db 421 IAGILPDLLVKGDDYKPEETAGSKEYWANGGEVLVNFEDGCSSTTNIKKIQDCKG 477
RESULT 2
US-09-912-020-325
; Sequence 325, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 325
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-325

Query Match 100.0%; Score 2393; DB 10; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.7e-191;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVWVGDVMDRYWYGTPTSRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60
Db 1 MKVTLPEFERAGVWVGDVMDRYWYGTPTSRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60
QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTITKLRVLSRNOQLRLD 120
Db 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTITKLRVLSRNOQLRLD 120
QY 121 FEFGFEGVDPQPLHERINQALSSIGALVLSYAKGALASVQOQMLARKAGVPLIDPKG 180
Db 121 FEFGFEGVDPQPLHERINQALSSIGALVLSYAKGALASVQOQMLARKAGVPLIDPKG 180
QY 181 TDFERYRGATLLPNLSEFEAVVKGCKTEEEIIVERGKMLIADYELSALLVTRSEOGMSLL 240
Db 181 TDFERYRGATLLPNLSEFEAVVKGCKTEEEIIVERGKMLIADYELSALLVTRSEOGMSLL 240
QY 241 QPKAPLHMPDQAEVYDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVGKLG 300
Db 241 QPKAPLHMPDQAEVYDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVGKLG 300
QY 301 STVSPTELENAVGRADTGFGVMTTEELKLAAGNSLEEACFFANAAAGVVGKLG 360

Db 301 STVSPTELENAVGRADTGFGVMTTEELKLAAGNSLEEACFFANAAAGVVGKLG 360
QY 361 ANARKGLDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALAEVDVWVVSFEEDTPQRL 420
Db 361 ANARKGLDRLIVAVNSDASTKRLKGDSPVNPVLEQRMIVLGALAEVDVWVVSFEEDTPQRL 420
QY 421 IAGILPDLLVKGDDYKPEETAGSKEYWANGGEVLVNFEDGCSSTTNIKKIQDCKG 477
Db 421 IAGILPDLLVKGDDYKPEETAGSKEYWANGGEVLVNFEDGCSSTTNIKKIQDCKG 477

RESULT 3

US-09-815-242-10301
; Sequence 10301, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10301
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10301

Query Match 100.0%; Score 2393; DB 10; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.7e-191;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVWVGDVMDRYWYGTPTSRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60
Db 1 MKVTLPEFERAGVWVGDVMDRYWYGTPTSRISPEAPVVPVVKVNTIEERPGGAANVAMNI 60
QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTITKLRVLSRNOQLRLD 120
Db 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTITKLRVLSRNOQLRLD 120
QY 121 FEFGFEGVDPQPLHERINQALSSIGALVLSYAKGALASVQOQMLARKAGVPLIDPKG 180
Db 121 FEFGFEGVDPQPLHERINQALSSIGALVLSYAKGALASVQOQMLARKAGVPLIDPKG 180
QY 181 TDFERYRGATLLPNLSEFEAVVKGCKTEEEIIVERGKMLIADYELSALLVTRSEOGMSLL 240
Db 181 TDFERYRGATLLPNLSEFEAVVKGCKTEEEIIVERGKMLIADYELSALLVTRSEOGMSLL 240
QY 241 QPKAPLHMPDQAEVYDVTGAGDTVIGVLAATLAAGNSLEEACFFANAAAGVVGKLG 300

```
Db 241 QPGKAPLHMPQAOEYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300
Qy 301 STVSPLEENAVRADTGFVMTTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
Db 301 STVSPLEENAVRADTGFVMTTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPPLQRMIVLGALEAVDWMVVSFEEDTPQRL 420
Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPPLQRMIVLGALEAVDWMVVSFEEDTPQRL 420
Qy 421 IAGILPDLVKGDDYKPEEIAAGSEKVEWANGGEVLVLPFEDGCGSTTNIKKIQDQKK 477
Db 421 IAGILPDLVKGDDYKPEEIAAGSEKVEWANGGEVLVLPFEDGCGSTTNIKKIQDQKK 477

RESULT 4
US-09-815-242-13759
; Sequence 13759, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13759
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13759
```

```
Query Match 94.7%; Score 2265; DB 10; Length 477;
Best Local Similarity 93.5%; Pred. No. 1.6e-180;
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;
```

```
Qy 1 MKVTLPEFERAGVMVGVDMVLDYWGPTSRISPEAPVPVVKVNTIERPGGAANVANNI 60
Db 1 MKVNLPAFERAGVMVGVDMVLDYWGPTSRISPEAPVPVVKVNTIERPGGAANVANNI 60
Qy 61 ASLGANARLVGLTGIDDAARALSLSLADNVNKCDFSVPTHTTITKRLVLSRNOOLIRLD 120
Db 61 ASLGANARLVGLTGIDDAARALSLSLADNVNKCDFSVPTHTTITKRLVLSRNOOLIRLD 120
Qy 121 FEEGFEVDPOPLHERINQALSSIGALVLSYAKGALASVQOMIQOLARKAGVPVLIDPKG 180
Db 121 FEEGFEVDPOPLHERINQALSSIGALVLSYAKGALASVQOMIQOLARKAGVPVLIDPKG 180
Qy 181 TDFERYRGATLLTNLSEFAVAGCKSEDELVERGMKLIADYLSALLVTRSEOGMTLL 240
```

```
Db 181 TDFERYRGATLLTNLSEFAVAGCKSEDELVERGMKLIADYLSALLVTRSEOGMTLL 240
Qy 241 QPGKAPLHMPQAOEYDVTGAGDTVIGVLAATLAAGNSLEECFFANAAAGVVGKLG 300
Db 241 QPNKAPLHMPQAOEYDVTGAGDTVIGVLAATLAAGNTLEACYFANAAGVVGKLG 300
Qy 301 STVSPLEENAVRADTGFVMTTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
Db 301 STVSPLEENAVRADTGFVMTTEELKLAFAAARKRGEKVVMTNGVFDILHAGHVSYL 360
Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPPLQRMIVLGALEAVDWMVVSFEEDTPQRL 420
Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPPLQRMIVLGALEAVDWMVVSFEEDTPQRL 420
Qy 421 IAGILPDLVKGDDYKPEEIAAGSEKVEWANGGEVLVLPFEDGCGSTTNIKKIQDQKK 476
Db 421 IAGILPDLVKGDDYKPEEIAAGSEKVEWANGGEVLVLPFEDGCGSTTNIKKIQDQKK 476

RESULT 5
US-09-815-242-12079
; Sequence 12079, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12079
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12079
```

```
Query Match 56.4%; Score 1350.5; DB 10; Length 474;
Best Local Similarity 57.9%; Pred. No. 1.4e-104;
Matches 274; Conservative 70; Mismatches 128; Indels 1; Gaps 1;
```

```
Qy 1 MKVTLPEFERAGVMVGVDMVLDYWGPTSRISPEAPVPVVKVNTIERPGGAANVANNI 60
Db 1 MKLSMPREDQAPVLVGVDMVLDYWGPTSRISPEAPVPVVRVEGHEDPGGAANVANNI 60
Qy 61 ASLGANARLVGLTGIDDAARALSLSLADNVNKCDFSVPTHTTITKRLVLSRNOOLIRLD 120
Db 61 AALGAQALLVGVGTGRDEADSLANSKAAAGVDAREQRIDTSQFTIVKLRVSRHOOLLKRD 120
Qy 121 FEEGFEVDPOPLHERINQALSSIGALVLSYAKGALASVQOMIQOLARKAGVPVLIDPKG 180
```

Db 121 FEEPER-TDAAALAVDVESSLAKVVLVLSYKYGKALONHQVLIQAARARNIPVLADPKG 179
QY 181 TDFERYGATLTLPNLSEFEAVVKGCKTEETIVBERGKMLIADYELSLVTRSEOGMSLL 240
Db 180 KDFATYRGASLTLPNLSEFEATVTRGCADEAEVLVAKGQALMSLDELGLLVTRGHEGHTLL 239
QY 241 QPGKAPLHMTQAOQYDVDTGAGDVIGVLAATLAAGNSLEACEFANAAAGVYVVGKLG 300
Db 240 RHGQALHLPARAREVDTGAGDVIVISTLAALAAAGEELPSAVGLANLAAGIVVGKLG 299
QY 301 STVSPLELAVNVRGADTGFVYMTTEELKLAVALAAARKGEKVVYVNTGVDILHAGHSYL 360
Db 300 AAISAPELRRAYVORQSGSERGVGLQEQLLLAIEDARAHGEKIVFTNGCFDILHAGHVTYL 359
QY 361 ANARKGLDRILVAVNSDASTKBLKGSRPVNPLEORMIVLGALEAVDWWVSEEDTPQRL 420
Db 360 EQARAQGDRLIYGVNDASVTRKGVGRPINSVDRMAVLGLGAVDWWVSAEDTPERL 419
QY 421 IAGILPDLLVKGDKPEIAGSKVWANGGEVVLVNFEDGCGSTTNIKKIQ 473
Db 420 LEQVRPDLVKGDDYGVGEVQVGAQIVKAYGGEVRVLGLVENSSTTAIVEKIRQ 472

RESULT 6

US-09-815-242-11854
; Sequence 11854, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11854

; LENGTH: 308

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-11854

Query Match 8.2%; Score 197; DB 10; Length 308;

Best Local Similarity 26.5%; Pred. No. 4.7e-09;

Matches 86; Conservative 50; Mismatches 153; Indels 36; Gaps 11;

QY 10 RAGVMYVGDVMDLRYWYGT-----SRISPEAPVPVVKVNTIERPFG--AANVAMNIA 63

Db 2 QAKVLVGSINMDLVVRAPRLPGGTLAGQSFITV-----PGKGANQAVAAARL 52

QY 64 GANARLVGLTGIDDAARALSKSLADVNVKDFVS-VPTHTTTKRLVLSRNOQLIRLDFE 122

Db 53 GAEVAMICCLGDGDYDQLYRALQAEGLDCQGVVERVAGESGVALIVVDDSSQNAIVVA 112
QY 123 EGFEGVDPOPL--HERINQALSSIGALVLSYAKGALASVOQMIQOLARKAGVPVLIDP-- 178
Db 113 GGNHLSFAVILARHHLLEQ-----AQVVCOLESFLETGVHVLRAHALGKTVILNPAP 167
QY 179 --KGDFERYRGATLTLPNLSEFEAVVKGCKTEETIVE---RGMKLIADYELSLVTR 233
Db 168 ATROVPAEWLPLVDYLVNPETESELL---CRLPVDSLESAGRAAERLRMGAGRVITLG 224
QY 234 EOGMSLLQPGKAPLHMTQAOQYDVDTGAGDVIGVLAATLAAGNSLEACEFANAAAGV 293
Db 225 AQGALLVGEGRVE-HFPVARVKALDTTAAGDTFVGFFAALARGLDEAAAIIRFGQAAAI 283
QY 294 VVGKLGSTVSPVPI--ELENVAVGRA 316
Db 284 SVTRUGAOTSIPSRREEVERALVGEA 308

RESULT 7

US-09-815-242-10892
; Sequence 10892, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10892

; LENGTH: 305

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-10892

Query Match 8.0%; Score 191.5; DB 10; Length 305;

Best Local Similarity 22.8%; Pred. No. 1.3e-08;

Matches 72; Conservative 63; Mismatches 142; Indels 39; Gaps 11;

QY 13 VMVVGDMLDRYWYGTSTRISPEAPVPVVKVNTIERPFG---AANVAMNIAISGANARL 69

Db 6 VTIIGSINDL-----TLRVK-EMPKGETIHAIEHFTAGGGKGANQAAAKSGAETVF 59

QY 70 VLTGTIDDAARALSKSLA--DVNVKDFVSVTPHTITKRLVLSRNOQLIRLD----- 120

Db 60 ICAGVNDGAGAMMTDLMSQDEINL-----TGVTTLEKTATGQAFIMVDNAGENSI 109

QY 121 --FEEGFEVDPPLPHERINQALSSIGALVLSDYAKGALASVQMIQOLARKAGVPVLIDP 178

```
Db 110 MIYAGANNAFTPKQVE--HQEIIIEKSDFVIAQF--ESAIDSTIAAFAKIAKAGVKTLINP 166
Qy 179 ---KGTDFERYRGATLLTNLSEFEAVVGVCKKTEEEIIVERGMKLIADYLSALLVTRSE 234
Db 167 APALEQVPEELLNVDMTVPNETETELTGKIIIDEASMRKAAEALHQLGIEAVIITVGS 226
Qy 235 QGMSLLQPGKAPLHMPHQAOQEVYDVTGAGDTVIGVLAATLAAG--NSLEEACFFANAAAGV 293
Db 227 KGAFYDVNGRSGI--VPAFKVAVDTTAAGDTFIGALSILEKDFSNLEEAIIRYGNKASSL 285
Qy 294 VGVKLGTSVSPIELE 309
Db 286 TVQRFGAQPSIPYQHE 301
```

RESULT 8

US-09-815-242-11035

Sequence 11035, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11035

LENGTH: 306

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-815-242-11035

Query Match 7.4%; Score 178; DB 10; Length 306;

Best Local Similarity 22.7%; Pred. No. 1.8e-07;

Matches 66; Conservative 45; Mismatches 116; Indels 64; Gaps 7;

```
Qy 51 GGAANVAMNTASLGANARLVGLTGIDDAARALSKSLADVNVKCDVSVPTHTTKLRVL 110
Db 40 KGKANQAAARLGAAGKAVFTSCIGSDSIGKTKNAFAQEGID-----TTH-----IN 86
Qy 111 SRNQLRLDFEEGFEVDPPLHERINQALSSIGALVLSYKAGLASVQOQMI-----164
Db 87 TVSQEMTGMAF-----IQVAKSSENSIVL---ASGANSLSSEMVVRQSEA 128
Qy 165 -----QLAKAGVPLVLDPKGTDF---ERYRGATLLTPNLSE 198
Db 129 QIAQSDCLLMQLETPLSGVLEAAQIAKNGVKVVLNPPAPAIQLSDELLSDIITPNTE 188
Qy 199 FEAVVGVCKTEEEIIVERGMKLIADYLSALLVTRSEQMSLLQPGKAPLHMPHQAOQEVYD 258
```

```
Db 189 AEILTGVEVADEQSAVRAASVFDKGIETVMITLGAKGVFVSRKGSRIIKGFCVQAI-D 247
Qy 259 VTGAGDIVIGVLAATLAAGNSLEEACCFANAAAGVGVVVKLGTSVSPIELE 309
Db 248 TTAAGDTFNGGFTVALLLEKSFDEAIRFGQAAAAISVTKKGAQSSIPTRQE 298
```

RESULT 9

US-09-741-669-479

Sequence 479, Application US/09741669

Patent No. US20020022718A1

GENERAL INFORMATION:

APPLICANT: Forsyth, R. Allyn

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

TITLE OF INVENTION: Genes identified as required for

TITLE OF INVENTION: proliferation of E. coli

FILE REFERENCE: ELITRA.009A

CURRENT APPLICATION NUMBER: US/09/741,669

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: US 60/173005

PRIOR FILING DATE: 1999-12-23

NUMBER OF SEQ ID NOS: 481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 479

LENGTH: 309

TYPE: PRT

ORGANISM: Escherichia coli

US-09-741-669-479

Query Match

Best Local Similarity 7.1%; Score 169; DB 10; Length 309;

Matches 77; Conservative 47; Mismatches 154; Indels 36; Gaps 9;

```
Qy 9 ERAGVMVY-GDVMLDRYWGPTSRISPEAPVYVVKVNTIEERPGS--AANVAMNTASLGAN 66
Db 2 QNAGSLVVLGSGINADHI----LNLQSFPTGETVTGNHYQVAFGKGKANQAAVAGRSCAN 57
Qy 67 ARVLGLTGIDDAARALSKSLADVNVKCDVSVPTHTTKLRVLSRNQOLRLDFEEGFE 126
Db 58 IAFIACTGDDSDIGESVROQLATDNI-----DITPVSVIKGESTGVALIFVNG-E 105
Qy 127 GVDPLHERINQALS-----SIGALVLSYKAGLASVQOQMIOLARKAGVPV 174
Db 106 GENVIGHAGANAALSPALVEAQERERANASALLMQ--LESPLSVMAAAKIAHONKTIV 163
Qy 175 LIDP---KGTDFERYRGATLLTPNLSEFEAVVGVCKTEEEIIVERGMKLIADYLSALLVT 231
Db 164 ALNPAPARELPDELLALVDIITPNTEAEKLTGIRVENDEDAKAAQVLEHEKGIRTVLIT 223
Qy 232 RSEGMSSLLQPGKAPLHMPHQAOQEVYDVTGAGDTVIGVLAATLAAGNSLEEACCFANAAA 291
Db 224 LGSRGVWASVNGEQ--RVPGFRVQAVDTIAAGDTFNGALITALLLEKPLPEAIRFAHAA 282
Qy 292 GVVGVKLGTSVSP 305
Db 283 AIAVTRKGAQPSVP 296
```

RESULT 10

US-09-815-242-10406

Sequence 10406, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard


```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13883
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Salmonella typhi
; US-09-815-242-13883

Query Match      6.9%; Score 165; DB 10; Length 309;
Best Local Similarity 24.8%; Pred. No. 2.2e-06;
Matches 67; Conservative 39; Mismatches 134; Indels 30; Gaps 6;

Qy 51 GGAANVAMNTIASGANARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTITKLRVL 110
Db 42 GKGANQAVAGRSGANTAFIACTGDDDDIGDSIRKQVSDRI-----DIAPVSVI 90

Qy 111 SRNQLRLDFEEFEGVDPOLHERINQALSSIGALV-----LSDYAKGALA 158
Db 91 KGESTGVALLIFVNG-EGENVIGIHAGANAALSP--ALVDAQRIARAQADALLMOLESPL 147

Qy 159 SVQMIOLARKAGVPLIDP---RGTFERYRGATLLTPNLSEFEAVVGVCKCKTEEEIVER 215
Db 148 SYLAANKIAHQHHTVALNPAPARELPDELLALVDIITPNETAEKLTGVRENDDDVAK 207

Qy 216 GMKLIADYELSAALVTRSEQMSLLQPKAPLHMTQAQEVYDVTGAGDTVIGVLAATLA 275
Db 208 AAQVLHDKGIRTVLTILGSRGWASVNGEG--RRVPGFSVEAVDTIAAGDTENGAFITAIL 266

Qy 276 AGNSLEECFANAAAGVVVVKLGTSVSP 305
Db 267 EETPLPAIRFAHAAAIAIVTRKGAQPSVP 296

RESULT 13
US-09-815-242-10180
; Sequence 10180, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10180
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10180

Query Match      6.6%; Score 158; DB 10; Length 309;
Best Local Similarity 25.4%; Pred. No. 8.2e-06;
Matches 80; Conservative 37; Mismatches 122; Indels 76; Gaps 14;

Qy 29 TSRSPE-----APVPVVKVNTTEERPGGAANVAMNTIASGANARLV-----GLTGIDDA 78
Db 19 TPQIYPEGKLRCTAPV-----FEPGGGINVARAIAHLGGSAIAIFPAGGATG----- 66

Qy 79 ARALSKSLADVNVKDFVSVPTHTITKLRVLSRNOQLRLDFE---EGFEGVDP----- 130
Db 67 -EHLVSLADENV-----PVATVEAKDTRQNLHVHVEASGEQYRFVHPGAALN 114

Qy 131 ----QPLHERINQALSSIGALVLSDYAKGA-LASVQOQMIOLARKAGVPLIDPKGTDFER 185
Db 115 EDEFROLEEQVLETESGAILVISGLSPLPGVKLEKLTQLISAAOKGIRCIIVSSG---EA 171

Qy 186 YRGA-----TLLTPNLSEFEAVVVK-----CKTEEEIVERG--MKLIADYELSALL 229
Db 172 LSAALAIIGNIELVLPNOKELSALVNRELTPDDVRKAAQAEIVNSGKARVVVSLGPGQAL 231

Qy 230 VTRSEQMSLLQPKAPLHMTQAQEVYDVTGAGDTVIGVLAATLAAGNSLEECFFRANA 289
Db 232 GVDSENCIQVPP-----PVKSO---STVGAGDSVGVGANTLKLAEASLEEMVRFVGA 281

Qy 290 AAGVVVVKLGTSVTS 304
Db 282 AGSAATLNOGTRLCS 296

RESULT 14
US-09-815-242-5399
; Sequence 5399, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```


; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5399
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5399

Query Match
Best Local Similarity 6.1%; Score 145; DB 10; Length 132;
Matches 45; Conservative 29; Mismatches 48; Indels 20; Gaps 6;
QY 341 KVMVTNGVFDILHAGHVSFLANARKLDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVL 400
DB 2 KRVTYGTVDLLHYGHIELLRAREMGDYLIIVALSTD-EFNQIK-HKKSYYDYEQKXML 59
QY 401 GALEAVDWV-----SFEEDTPQRLTAGILPDLLVKGDDYKPEETAGSKEVWANGVELV 455
DB 60 ESIRYVDLVIPEKGWGQKEDDVEKEDV----DVFVGHWDWE-----GEFDLKDCKCEVIY 110
QY 456 LNFEDGCGSTTNIKKIQDQKKG 477
DB 111 LKRTGISTT-----KIKQELYG 128

RESULT 15

US-09-815-242-12303
; Sequence 12303, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12303
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12303

Query Match
Best Local Similarity 6.1%; Score 145; DB 10; Length 132;
Matches 45; Conservative 29; Mismatches 48; Indels 20; Gaps 6;
QY 341 KVMVTNGVFDILHAGHVSFLANARKLDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVL 400
DB 2 KRVTYGTVDLLHYGHIELLRAREMGDYLIIVALSTD-EFNQIK-HKKSYYDYEQKXML 59
QY 401 GALEAVDWV-----SFEEDTPQRLTAGILPDLLVKGDDYKPEETAGSKEVWANGVELV 455

DB 60 ESIRYVDLVIPEKGWGQKEDDVEKEDV----DVFVGHWDWE-----GEFDLKDCKCEVIY 110
QY 456 LNFEDGCGSTTNIKKIQDQKKG 477
DB 111 LKRTGISTT-----KIKQELYG 128
Search completed: November 24, 2002, 23:21:07
Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run On: November 26, 2002, 10:19:07 : Search time 298 Seconds
(without alignments)
3604.709 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLPEFERGVVGVDM.....FEDGCTTIKIQDQKKG 477

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n_model -DEV=xlp
-O=/cgn2_1/USPTC_spool/US09912020/runat_25112002_091428_24237/app_query.fasta_1.647
-DB=N_Geneseq_101002 -QFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -DOORCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09912020.ecgn_1.1.281_etrunat_25112002_091428_24237 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMSOUP=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2393	100.0	1434	21	AAA65973	E. coli proliferat
2	2393	100.0	1434	22	AAH81477	Escherichia coli p
3	2393	100.0	1434	23	AAS52567	E. coli DNA for ce
4	2265	94.7	1434	23	AAS56024	Salmonella typhi D
5	1712	71.5	1830121	17	AAH42063	Haemophilus influe
C	1350.5	56.4	1425	23	AAS54344	Pseudomonas aerugi
C	1036	43.3	1001	22	AAF91438	Haemophilus influe
C	1036	43.3	1001	24	ABK37818	DNA sequence upstr
9	856	35.8	707	24	ABQ21090	Oligonucleotide fo
C	856	35.8	707	24	ABQ21091	Oligonucleotide fo
C	753.5	31.5	78845	21	AAH81463	N. meningitidis pa
12	753.5	31.5	349880	21	AAF21608	Neisseria meningit
13	745.5	31.2	969	24	AAS97221	Neisseria meningit
14	734.5	30.7	1272	18	AAT68016	H. pylori cytoplas
C	696	29.1	707	24	ABQ21092	Oligonucleotide fo
16	696	29.1	707	24	ABQ21093	Oligonucleotide fo
17	683	28.5	2055	23	AAS89669	DNA encoding novel
C	637	26.6	555	24	ABQ45444	Oligonucleotide fo
19	610.5	25.5	58857	21	AAA58471	Nucleotide sequenc
C	551	23.0	555	24	ABQ45442	Oligonucleotide fo
22	551	23.0	555	24	ABQ45443	Oligonucleotide fo
C	506.5	21.2	3552	23	AAS73093	DNA encoding novel
24	498.5	20.8	1344	23	AAS88908	DNA encoding novel
25	471.5	15.7	706	24	ABQ20706	Oligonucleotide fo
C	471.5	15.7	706	24	ABQ20707	Oligonucleotide fo
28	348.5	14.6	13574	21	AAH1529	N. meningitidis pa
C	348.5	14.6	173325	21	AAF21613	Neisseria meningit
C	348.5	14.6	837096	21	AAH1489	N. meningitidis pa
C	328.5	13.7	706	24	ABQ20708	Oligonucleotide fo
31	328.5	13.7	706	24	ABQ20709	Oligonucleotide fo
32	201	8.4	679	24	ABQ38428	Oligonucleotide fo
C	201	8.4	679	24	ABQ38429	Oligonucleotide fo
34	197	8.2	927	23	AAS54120	Oligonucleotide fo
C	192	8.0	34980	22	AAH41226	Pseudomonas aerugi
36	191.5	8.0	918	23	AAS53158	Pyrococcus abyssi
37	190.5	8.0	1283	20	AAH13469	Enterococcus faeca
C	182.5	7.6	6948	20	AAH12941	Enterococcus faeca
39	181.5	7.6	880	24	ABK73456	Bacillus lichenifo
40	181.5	7.6	888	24	ABK73548	Bacillus lichenifo
41	178	7.4	921	23	AAS53301	Haemophilus influe
42	178	7.4	1830121	17	AAH42063	Haemophilus influe
43	172	7.2	1314	23	AAS82459	DNA encoding novel
44	170	7.1	243	24	ABN77217	Human ORF2164 cDNA
45	169	7.1	930	22	AAH81487	Escherichia coli p

ALIGNMENTS

RESULT 1
AAA65973
ID AAA65973 standard; DNA; 1434 BP.
XX
AC AAA65973;
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated coding sequence SEQ ID NO:165.
XX
KW Escherichia coli; E. coli; proliferation; inhibition; screening;
XX antimicrobial; Bacterial growth; antisense therapy; antibacterial; ds.
OS Escherichia coli.
XX
PN WO200044906-A2.
XX
PD 03-AUG-2000.

XX 27-JAN-2000; 2000WO-US02200.
 XX
 XX 27-JAN-1999; 99US-0117405.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
 XX Yamamoto RT, Xu HH;
 XX WPI; 2000-514822/46.
 XX P-PSDB; AAB15968.
 XX
 XX Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 XX
 XX Claim 8; Page 146-147; 316pp; English.
 XX
 XX AAA65809 to AAA65989 and AAA66058 to AAA66138 represent nucleotide
 CC sequences derived from *Escherichia coli* which inhibit *E. coli*
 CC proliferation. AAA65989 to AAA66055 and AAB15886 to AAB16040 represent
 CC nucleotide and protein sequences associated with *E. coli* proliferation.
 CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods from the present invention can be used to
 CC identify a proliferation- required gene in a microorganism, by contacting
 CC a microorganism with a proliferation- required gene activity inhibitory
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences
 CC identified as being required for bacterial growth and proliferation, can
 CC be used for antisense therapy for killing bacteria.
 XX
 XX Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 2,92e-211 Length: 1434
 Score: 2393.00 Matches: 477
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-912-020-325 (1-477) x AAA65973 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
 DB 1 ATGAAAGTAACGCTGCCAGAGTTTGAACGTGCAGGAGTGATGGTGGTGTGATG 60
 QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
 DB 61 CTGGATCGTTACTGGTACGGCCGCCACAGTCGTATCTCGCGGAGCGCGGTGCCG 120
 QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
 DB 121 GTTAAAGTGAAATACCATCGAAGAACGTCCTCGGGCGCGGCTAACGTGGCGGATATC 180
 QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
 DB 181 GCTTCTCTCGGTGCTAATGACGCGCTGGTGGGTGTGACGGGCAATTGACGATGACGCGC 240
 QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
 DB 241 GCGCTGAGTAATCTCTGGCCGAGCTCAAGTCAATGCGACTTCGTTCTGTACCGACG 300
 QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
 DB 301 CATCCGACCATTAACCAATACGGGTACTTTCGCGCAACCAACAGCTGATCCGCTCGAT 360
 QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
 DB 361 TTGGAAGAGGTTTCGAAGGTGTGTGATCCGCGAGCCCTGTCACGAGCGGATTAATCAGCGC 420

QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
 DB 421 CTCAGTTTCGATTGGCGCGCTGGTGTCTTCTGACTACGCCAAAGGTGGCTGGCAAGCGTA 480
 QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
 DB 481 CACGACATGATCAACTGGCGCGTAAAGCGGGTGTCTCCGGTGTGATTGATTCACAAAGGT 540
 QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
 DB 541 ACCGATTTTGAGCGCTACCGCGCGCTACCGCTGTGTACGCCGAATCTCTCGGAATTTGAA 600
 QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
 DB 601 GCTGTTGCTCGTAATGTAAGACCGAAGAGAGATTGTTGAGCGCGSCATGAATGATT 660
 QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
 DB 661 GCGGATTACGAACCTCGGCTCTGTAGTGACCGCTTCCGAACGAGGTATGTGCTGCTG 720
 QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
 DB 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCAGCGGAGAGTGTATGACGTTACC 780
 QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
 DB 781 GGTGCGGGCGACACGCTGATTGGCGTCTCGCGCAACGCTGGCAGCGGGTAAATTCGCTG 840
 QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
 DB 841 GAAGAAACCTGCTTCTTTGCAATGCGCGCGCTGGTGGTGGTTCGCGCAACATGGGAACC 900
 QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
 DB 901 TCCACGGTTTCGCGGATCGAGCTGGAAATGCTGTAGCTGGAGCTGCAGATACAGGCTTT 960
 QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaAlaArgLysArgGlyGlu 340
 DB 961 GCGCTGATGACCGAAGAGAACTGAAGCTGGCGTAGCGCGCAGCGCTAAACGTGGTGAA 1020
 QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
 DB 1021 AAAGTGGTGTATGACCAACGCTGCTTTGACATCTTCACCGCGCGGACGCTCTCTATCTG 1080
 QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
 DB 1081 GCAATGCCCGCAGCTGGGTGACCGCTTGATGTGTCCTCAACACGATGCTCCACCC 1140
 QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
 DB 1141 AAACGGCTGAAGGGGATTCGCCCGGTAAACCCACTCGAACACGCTATGATTGTGCTG 1200
 QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
 DB 1201 GCGCGCACTGGAAGCGGTGCGATGCTGGTGTGCTGTTTGAAGAGGACACCGCGCAGCTG 1260
 QY 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyAspTyrLysProGluGluIle 440
 DB 1261 ATGCCGGGATCTTCCAGATCTGCTGGTGAAGCGCGGCGACTATAAACCAAGAGAGATT 1320
 QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 DB 1321 GCCGGGAGTAAAGAGTCTGGGCCAACCGTGGGCAAGTGTGGTGTCTCAACTTTGAAGAC 1380
 QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLysGly 477
 DB 1381 GGTGTGCTGCGACCAACATCATCAAGAAGATCCAACAGGATAAAAAAGGC 1431

RESULT 2
 AAH81477
 ID AAH81477 standard; DNA; 1434 BP.
 XX
 AC AAH81477;

```

XX 21-SEP-2001 (first entry)
XX Escherichia coli protein encoding nucleotide sequence SEQ ID NO:276.
XX
XX Escherichia coli; identification; proliferation; microorganism;
XX antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX bacterial growth inhibition; ds.
XX
XX Escherichia coli.
XX
XX WO200148209-A2.
XX
XX 05-JUL-2001.
XX
XX 19-DEC-2000; 2000WO-US34419.
XX
XX 23-DEC-1999; 99US-0173005.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Forsyth RA, Ohlsen KL, Zyskind JW;
XX
XX WPI; 2001-457376/49.
XX
XX P-PSDB; AAG98421.
XX
XX Novel nucleic acids encoding proteins required for Escherichia coli
XX proliferation, useful for screening for antimicrobial agents -
XX
XX Example 4; Page 415-417; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
XX sequence (I) consisting essentially of one of the 93 nucleotide sequences
XX given in AAH81202 to AAH81294, where expression of the nucleic acid in a
XX microorganism is capable of inhibiting proliferation of a microorganism.
XX (I) have antibacterial and antibiotic activities, and can be used in
XX gene therapy. Expression of (I) in a microorganism inhibits proliferation
XX of the microorganism, and the manufactured antibiotic is useful for
XX reducing the activity or level of a gene product required for
XX proliferation of a microorganism in a subject, specifically humans. The
XX nucleic acids that inhibit bacterial growth or proliferation can be used
XX as antisense therapeutics for killing bacteria. In addition to
XX therapeutic applications, the nucleic acid sequences complementary to
XX sequences required for proliferation can be used as diagnostic tools.
XX For example, nucleic acid probes complementary to proliferation-required
XX sequences that are specific for particular species of microorganisms can
XX be used as probes to identify particular microorganism species in
XX clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
XX proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
XX represent oligonucleotides, which are used in the exemplification of the
XX present invention.
XX
XX Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,92e-211 Length: 1434
XX Score: 2393.00 Matches: 477
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-912-020-325 (1-477) x AAH81477 (1-1434)
XX
XX 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
XX
XX 1 ATGAAGTAACCGTCCAGAGCTTTGACCGTGCAGCGATGATGGTGGTGGTGGTGGTGGT 60
XX
XX 21 LeuAspArgTyrTyrTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
XX
XX 61 CTGGATCGTTACTGGTACGGCCGCCACCAGTGTATCTCGCGGAAGCCGCGTGGTGGT 120
XX
XX 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60

```

```

Db 121 GTTAAAGTGAATACCATCGAAGAACGCTCCGGCGCGCGGCTACGTCGCCATGAATATC 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCTTCTCTCGGTGCTAATGACGCGCTGGTCGGGTGACGGCGCATGACGATGACGCGGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 CGCTGAGTAATCTCTGGCCGACGTCAACGTCAATCGGACTTCGTTTCTGTACCGACG 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATACCAATACGGGTACTTTCCGCAACCAACAGCTCATCGTCTGGAT 360
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAAGGTTTCGAAGGTGTTGATCCGCGCGCTGACGAGCGGATTAATCAGCGC 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTGATGGCGCGCTGGTCTTCTGACTACGCCAAAGGTGCGCTGGCAAGCGTA 480
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGCAGATGATCCAACTGGCGGTAAAGCGGTGTTCCGGTCTGCTGATTGATCCAAAGGT 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGAGCGCTACCGCGCGCTACGCTGTAAAGCGCGAACTCTCCGNAATTTGAA 600
Qy 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTTGTCGTAATGTAAACCGAAGAGAGATGTTGAGCGCGGATGAAACTGATT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCCGATTACGAACCTCTCGGCTCTGTAGTACCGCTTCCGAACAGGATGATGCGTGGTG 720
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnGlnGlnValValTyrAspValThr 260
Db 721 CAACCGGTAAAGCGCGCTGATATGCCAACCCAGCGAGGAGTGTATGACGTTTACC 780
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGCGCACACGGTATTGGCGTCTCGCGGCAACGCTGGCAGCGGGTAATTCGCTG 840
Qy 281 GluGluAlaCysPheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAAGAAGCCTGCTTCTTTGCCAATCGCGCGCTGGCGTGGTGGTGGCAAACTGGGAACC 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTTCGCCGATCGAGCTGGAAATGCTGTACGTGGCGGTGACAGATACAGGCTTT 960
Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTGATGACCGAGAGAGAACTGAGCTGGCGGTAGCGGCGGCGGTAAACGTTGTA 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAGTGGTGATGACCAACGGTCTCTTTGACATCTCGACGCGCGGCGGCGTCTCTTATCTG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAATGCGCGCAAGCTGGGTGACGCTGTGTTGTCGCTCAACAGCGGATGCTCCACC 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGGCTGAAAGGGGATTCGCGCGGTTAAACCCACCTCGAACAGCGGTATGATTGTGCTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420

```

Db 1201 GCGCCTGGAAGCGCTGACCTGGGTAGTGTCTTTGAAGAGGACACGCCGCGCTTG 1260
QY 421 IIEAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGluGluIle 440
Db 1261 ATGCCGGGATCTGCCAGATCTGCTGGTGAAGCGCGGCTATATAACCAAGAGATT 1320
QY 441 AlaGlySerLysGluValTyrAlaAsnGlyGlyGluValLeuValLeuAsnGluAsp 460
Db 1321 GCGGGAGTAAGAAGTCTGGGCCACAGCTGGCGAAGTGTGGTCTCACTTTGAGAC 1380
QY 461 GlyCysSerThrThrAsnIleLeuLysLysIleGlnGlnAspLysLysGly 477
Db 1381 GGTGTCTGAGGACCAACATCATCAAGAAGATCAACAGGATAAAAAAGGC 1431

RESULT 3

AAS52567

ID AAS52567 standard; DNA; 1434 BP.

XX

AC AAS52567;

XX

DT 13-FEB-2002 (first entry)

XX

DE E. coli DNA for cellular proliferation protein #289.

XX

KW Antisense; ds; prokaryotic cellular proliferation gene;

XX

KW antibiotic; antibacterial; drug design.

XX

OS Escherichia coli.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US09180.

XX

PR 21-MAR-2000; 2000US-191078P.

XX

PR 23-MAY-2000; 2000US-206848P.

XX

PR 26-MAY-2000; 2000US-207727P.

XX

PR 23-OCT-2000; 2000US-242578P.

XX

PR 27-NOV-2000; 2000US-253625P.

XX

PR 22-DEC-2000; 2000US-257931P.

XX

PR 16-FEB-2001; 2001US-269308P.

XX

(ELIT-) ELITRA PHARM INC.

XX

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX

PI Yamamoto RT, Xu HH;

XX

PI WPI; 2001-611495/70.

XX

DR P-PSDB; AAU34708.

XX

DR New polynucleotides for the identification and development of

XX

PT antibiotics, comprise sequences of antisense nucleic acids -

XX

PS Claim 27; Seq ID No 6204; 51lpp; English.

XX

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic

CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 other;

Alignment Scores:

Pred. No.: 2,92e-211 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-912-020-325 (1-477) x AAS52567 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAAGTAACGCTGCAGAGTTTGAACCTGCAGGAGTATGCTGTTGTTGATGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTTACTGGTACGCCCCACCAGTCGTATCTCCCGGAAGCGCGGTGCCG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTCAATACCATCGAAGACCTCCGGGGCGCGCTAACCTGGCGATGAATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCTTCTCTCGGTCTAATGCAGCCTGTTGGTGGGTTTACGGGCATTGACGATGACGCGC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGAGTAATATCTCTGGCGACGTCACAGTCAATGACGACTTCGTTCTGTACCGACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATTAACCAATTAACGGGTACTTCCCGCAACCAACAGCTGATCCTCTGGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAGGTTTCGAAGGTGTTGATCCGACGCCCTGCACGAGCGGATTAATACGGCG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTCAGTTCGATTGGCGCGCTGGTGTCTTCTGACTACGCCAAAGCTCGCTGGCAAGGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGCAGATGATCCAACTGGCGCTTAAAGCGGTGTTCCCGTGTCTGATTGATCCAAAGGT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGGAGCGCTACCGCGCGCTACGCTGTTAACCGCGAATCTCTCGGAATTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValIleAlaArgGlyMetLysLeuIle 220
Db 601 GCTGTTGTCGGTAAATGTAAGACCGAAGAGAGATTGTTGAGCGCGCATGAACATGATT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCGGATTAGCACTCTCGGCTCTGTTAGTACCCGTTCCGAACACAGGATATGTCGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAAACCGGGTAAAGCGCGCTGCATATGCCAACCCAGCGAGGAAGTGTATGACGTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGCGACACGCTGATTGGCGTCTTGGCGCAACGCTGCGCAGCGGTAATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValGlyLysLeuGlyThr 300

Db 841 GAAGAAGCCTGCTTCTTTGCAATCGCGCGCTGCTGCTGCGCAAACTGGGAACC 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTGCGCGATCGAGCTGGAATGCTAGTGCAGCTGACATACAGGCTT 960
Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTGATGACCAAGAGGAACCTCAAGCTGCGCGTAGCGCGGTAAACGTGGTAA 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAATGGTGATGACCAACGGTGTCTTTGACATCTGACGCGCGGCGCTCTTATCTG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAATGCCCGCAAGCTGGGTGACCGCTTGATTTGCGCTCAACAGCGATGCTCCACC 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AACCGGTGAAGGGGATTCGCGCGGTAAACCCACTCGAAGAGCGTATGATTGTGCTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCCTGGAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATCCCGGGATCTGCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 441 AlaGlySerLysGluValTyrPheAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCCGGGAGTAAAGAGCTGCGGCAACGCTGGCGAAGTGTGGTGTCTCAACTTTGAAGAC 1380
Qy 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnAspLysLysGly 477
Db 1381 GGTTCGCTCGAGCACCATCATCAAGAAGATCCCAACAGGATAAAAAAGGC 1431

RESULT 4
AAS56024
ID AAS56024 standard; DNA; 1434 BP.
XX AAS56024;
AC AAS56024;
AT 13-FEB-2002 (first entry)
DT Salmonella typhi DNA for cellular proliferation protein #57.
XX Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX Salmonella typhi.
OS Salmonella typhi.
PN WO200170955-A2.
XX 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.

DR P-PSDB; AAU38165.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX Claim 27; Seq ID No 9661; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1434 BP; 311 A; 360 C; 468 G; 295 T; 0 other;
Alignment Scores:
Pred. No.: 1.88e-199 Length: 1434
Score: 2265.00 Matches: 445
Percent Similarity: 97.48% Conservative: 19
Best Local Similarity: 93.49% Mismatches: 12
Query Match: 94.65% Indels: 0
DB: 23 Gaps: 0
US-09-912-020-325 (1-477) x AAS56024 (1-1434)
Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTAAATCTGCCAGCGCTTTGAACGTCGACGCGCTCATGTTGGTGGTAAATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGATCGCTATGTTGGTATGCCCGCCACTTGCCTATTTCCACCGAAGCGCGGTGCCGTG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAGGTAAATACCGTTGAGGAACGTCGCGGCGCGCGGGAACGTCGCGCATGAACATT 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCGTCTCTGGGAGGAACGCGCGCTGCTGCTGCGCTGACGGGTATGATGACGCGCGCGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGACAAACGCTGGCGGAGGTCAATGTGAAGTCGACTTCGTTTCTGTCGCGACG 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACGATTACAAACTCGCGGTACTATACGTAATACGACAGCTCATTCGTCCTGAT 360
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAG3CTTTGAGGGCGTGACCGCGCGCTGTCATGAGCGGTATCAACGACGCG 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGGGATGATCGCGCGCTGCTGTTTCCGATTTATGCAAGGGCGCTCTCACCAGCGTG 480
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180

481 CAGACTATGATTCCTAGCGGCCAGCGCGCGTGCCTCATCATCCGAAGA 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACGGATTTGAACGTTACCGCGCGCCACGCTGCTGACGCCAAACCTTCTGAATTCGAG 600
Qy 201 AlaValValGlyLysCysLysThrGluGluGluLeuValGluArgGlyMetLysLeu 220
Db 601 CGCGTTGCGGGAAATGTAAGCAAGCAAGCAACTGGTTGAACGCGCATGAAGCTCAT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeu 240
Db 661 GCGGATTTAGACCTTTCGCGCTGTTGGTACGCGCTTCGGAACAGGAATGACGCTG 720
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGAATAAGCGCGCTACATATGCGACGCGAGCGCGCAAGTATATGATGTTACC 780
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGGATACGGTATCGCTGCTGCGCGACGCTGGCGCGGAAATACCTG 840
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
Db 841 GAGGAGCGGTGTTATTCGCCAATGCGCGCGCGCTAGTGGTAAACCTCGGAG 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCAACGGTTTCCTTATTCAGCTGGAAACGCGAGTGGCGGACGCGGATACCGGCTC 960
Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGCTCATGCGGAAGAGAGTGTGACAGCGCGCTGCGCGCGCGTAAAGCGTGGCGAG 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAGTGTCTATGACCAACGGCGTTTCGATATCTGCACGCGCGGCGACGCTCTTATG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAlaAsnSerAspAlaSerThr 380
Db 1081 GCGAAGCGCGCAAACTGGCGACCGCTGATTGTTGGCGTCAATAGTACGCGCTCGACT 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGCTCTAAAGGCGAAAGCGCTCGGTTAATCCGCTCGAAGCGGATGATCGTGTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCGCTCGAGTGGTCTGCTGCTGCTCTTTTGAAGAGGATACGCGCGCAACGACTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyLysProGluGluIle 440
Db 1261 ARTGCGGTTATTCGCGGATCTGCTGGTAAAGGCGGCGGACTATAAGCGGAGAGATC 1320
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCGGCGAGCGAAGAGCTGCGCAACGCGCGGCGAAGTATGCTGCTGAACCTCGAAGAT 1380
Qy 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLys 476
Db 1381 GGTGTTCCACCAACCAATATCATCAAAAAGATCCAGACCGGAGCGGAG 1428
RESULT 5
AAT42063/C
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;

expression modulating fragment; regulation; gene expression; vector;
organism; open reading frame; ORF; ds.
Haemophilus influenzae.
WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UJJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:
Pred. No.: 2.09e-144 Length: 1830121
Score: 1712.00 Matches: 331
Percent Similarity: 83.48% Conservative: 58
Best Local Similarity: 71.03% Mismatches: 77
Query Match: 71.54% Indels: 0
DB: 17 Gaps: 0
US-09-912-020-325 (1-477) x AAT42063 (1-1830121)
Qy 7 GluPheGluArgAlaGlyValMetValValGlyAspValMetLeuAspArgTyrTrpTyr 26
Db 1596552 GAATTTAAGCAAGCAAAAGTACTCGTATAGCGACGTGATCGTGTATTTGGTTC 1596493
Qy 27 GlyProThrSerArgLysSerProGluAlaProValProValValLysValAsnThrIle 46
Db 1596492 GCGCAACCAACCGTATTTCCACGAGACACCAGTACGTTGCTGTCGTACAGAAAT 1596433
Qy 47 GluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn 66
Db 1596432 GAAGAACGCGGGTGGTGCAGCAAAATGGCGATGAATATTGCTTCACTCAATGTACCC 1596373
Qy 67 AlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeu 86
Db 1596372 GTTCAGTTAATGGGATGATTGGACAAAGATGAAACTGGTTCTGCACCTTCCCTCTTATA 1596313
Qy 87 AlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLys 106
Db 1596312 GAAAAACAAAAAATGATTGTAATTTGTCATTAGAAACCCATCCAAACCATTTACTAAA 1596253
Qy 107 LeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyPheGlu 126

US-09-912-020-325 (1-477) x AAS54344 (1-1425)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTTTGTCCATGCCCGCTTCGACAGCCCGCGTGTGGTGTGGTGTGGTGTGGT 60
QY 21 LeuAspArgTyrTriPyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTCACCGCTACTGGCATGGCGGACTTCGGGCATTTCCCGGAGGCCCGCGGTGGCGGTG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTCCGGCTGCAACAGCAGCAGGACCGCCCGCGCGCGCCCAACGTCGCGTGAACATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCGCGCTGGGCGCGCAGGCGCTTGTGGTGGCGTCAACCGCGCGCAGCAGGCGCGCGAC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 AGCCTGGCCCAACAGCCTCAAGCGCGTGGAGTGGACGCGCGCTTCCAGCGCATCGATACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CAGCCGACCATCTCAAGCTGGGGTTCATGAGTCGCCACCACTGCTGGCGGTGCGAC 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLysHisGluArgIleAsnGlnAla 140
Db 361 TTCGAGAACCGTTCCG 417
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 418 CTGCGCAAGTCAAGTGTGGTGTGCTCCGACTACGCGCAAGCGCGCGCTACAGAACCCAC 477
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 478 CAGGTGCTGATCAGCGCGCGCGCGCGCGCAACATTCGCGTACTGCGCGATCCCAAGGCG 537
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 538 AAGACTTCGCCATCATTCG 597
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 598 ACCATGCTGGCGGTTCGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 658 AGCGAACTCGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnValValThrAspValThr 260
Db 718 CGCCATGGCG 777
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 778 GGTGGCGCGGATACGTCATCCACCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 838 CCTTCGCGCGGTGGTGTGGCAACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 898 GCGGCGCTAGCG 957
QY 321 GlyValMetThrGluGluGluLysLeuAlaValAlaAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 958 GCGCTGCTGGCGCTGAGCAATTCGCTGGCAATCGACAGACCGCGCGCGCGCGCGCG 1017
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
|||||

Db 1018 AAGATCGTCTTCCACCAATGGCTGCTTCGACATCCTTACGCCCGCCACGCTACCTC 1077
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1078 GAACAGCG 1137
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1138 ACTCGCTCAAGGGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1198 GCGGGCTGCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1257
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1258 CTCGACGAGTGGTCCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1317
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1318 GTCGGCGCGCAGATCGTCAAGGCTACGCGCGCGGAGGTACGGGTGCTCGCGCTGGTGG 1377
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGln 473
Db 1378 AACAGTCCACCACCGCCATCGTCTGAGAAATCCGCCAG 1416

RESULT 7
AAF91438/C
ID AAF91438 standard; DNA; 1001 BP.
XX
AC AAF91438;
XX
DT 04-MAY-2001 (first entry)
XX
DE Haemophilus influenzae (H1rd) HtrB gene upstream sequence, SEQ ID:64.
XX
KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PS (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Claim 46; Page 93-94; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also

CC encompasses modified Gram-negative bacterial strains from which the bleb
 CC preparations are made, a vector suitable for performing recombination
 CC events (for the generation of the modified bacterial strains),
 CC bacterially-derived nucleic acid sequences used in such a vector, and an
 CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
 CC cell vaccine suitable for paediatric use. The bleb preparation is useful
 CC in the manufacture of a medicament for immunising a human host against a
 CC disease caused by infection of one or more of the following: *Neisseria*
 CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Moraxella*
 CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*
 CC pneumonia. The invention may also be used to provide immunisation against
 CC the influenza virus. Bacterially derived nucleotide sequences of the
 CC invention are used in the performance of homologous recombination events
 CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
 CC increase or decrease expression of that gene. Immunoprotective and
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
 CC are more immunogenic, less toxic and safer, and are particularly useful
 CC for paediatric use. The present sequence represents a specifically
 CC claimed *Haemophilus influenzae* nucleic acid sequence.

XX
 SQ Sequence 1001 BP; 301 A; 225 C; 168 G; 307 T; 0 other;

Alignment Scores:

Pred. No.: 2,86e-86 Length: 1001
 Score: 1036.00 Matches: 201
 Percent Similarity: 82.37% Conservative: 42
 Best Local Similarity: 68.14% Mismatches: 52
 Query Match: 43.29% Indels: 0
 DB: 22 Gaps: 0

US-09-912-020-325 (1-477) x AAF91438 (1-1001)

Qy 7 GluPheGluArgAlaGlyValMetValValGlyAspValMetLeuAspArgTyrTrpTyr 26
 Db 886 GAATTTAAGCAAGCAAAAGTACTTGTATTAGCGCATGTGATGCTGTGATTTGGTTC 827
 Qy 27 GlyProThrSerArgIleSerProGluAlaProValProValValValValAsnThrIle 46
 Db 826 GGTCAACCAACCGTATTTCACCAAGAGCACCAGTCGAGTAGTTCGCGTACAGAAAT 767
 Qy 47 GluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn 66
 Db 766 GAAGACGCGCGCGGTGCTGCAAAATGTGGCGATGATATTTGCTTCACTCAATGTACCC 707
 Qy 67 AlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeu 86
 Db 706 GTTCAGTTAATGGATGATTGGACAGATGACAGAGTTCGACATTCACCTATTATA 647
 Qy 87 AlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLys 106
 Db 646 GAAAAACAAAGATTGATTGTAATTTTGTTCATTAGAAACCCATCCAACTTACTAAA 587
 Qy 107 LeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPheGlu 126
 Db 586 TTACGTATTTATTCGTCTCATCAACAGCTGCTCCGCTGATTTTGAAGAAGATTTCAT 527
 Qy 127 GlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAla 146
 Db 526 AATGTAGATTCAGAGATTATTAGCAAGTGTAGAAAGTCGGGTGAAAAATACGGTCT 467
 Qy 147 LeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeu 166
 Db 466 TTGATTTCTTCTGATTCGCGCAAGCGCGCTTAAAGATGTTCAAGAAATGATTCAAAT 407
 Qy 167 AlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArgTyr 186
 Db 406 GCAGCAAGCGCAATGTGCTGTGTTGATCGATCCAAAGGGAACGTATTTGAACGTTAT 347
 Qy 187 ArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCys 206
 Db 346 CGTGGCGGTACATTTATGACACCAATATGCTGAATTTGAAGCGGTGTAGGTAATGT 287
 Qy 207 LysThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSer 226

Db 286 AATACAGAGAGAGATATTAAAGAGTTTAAATTAATTTCTGATATTGAATTAACC 227
 Qy 227 AlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaPro 246
 Db 226 GCATTTTGTGTGACGCTCTGAAAAAGGCATGACATATTACGCCCAAAATCAAGAGCT 167
 Qy 247 LeuHisMetProThrGlnAlaGlnGluValThrAspValThrGlyAlaGlyAspThrVal 266
 Db 166 TATCATTTGCCAACCGTTGCAAAAGAGTGTGATGTGACGGGAGCGGTGACACTGTC 107
 Qy 267 IleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePhe 286
 Db 106 ATTAGTGTATTAGCAACCATAGCAGATGACGCTTCTTTTCAAGAAATCTGTGTACCTA 47
 Qy 287 AlaAsnAlaAlaAlaGlyValValValValGlyLysLeuGlyThrSer 301
 Db 46 GCCAATGTTGCGCAGGAATTTGTGTGTAATTTGGGACTTCA 2

RESULT 8

ABK37818/C

ID ABK37818 standard; DNA; 1001 BP.

XX AC ABK37818;

XX DT 08-MAY-2002 (first entry)

XX DE DNA sequence upstream of htrB #3 gene.

XX KW Upstream sequence; ds; Antibacterial; vaccine; bleb;

XX KW Gram-negative bacteria; outer membrane; LPS; lipopolysaccharide;

XX KW meningitis; bacteraemia; otitis media; pneumonia; chronic bronchitis;

XX KW sinusitis.

XX OS Haemophilus influenzae HIRd.

XX PN WO200209746-A2.

XX PD 07-FEB-2002.

XX PF 31-JUL-2001; 2001WO-EP08857.

XX PR 31-JUL-2000; 2000EP-0956369.

XX PR 08-FEB-2001; 2001GB-0003170.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Berthet FJ, Dalemans W, Denoel P, Dequesne G, Feron C, Garcon N;

XX PI Lobet Y, Poolman J, Thiry G, Thonnard J, Voet P;

XX DR WPI; 2002-188688/24.

XX PS New immunogenic composition comprising an antigen derived from a

XX PT pathogen and a bleb preparation from *Neisseria meningitidis*, useful as

XX PT a vaccine for treating or preventing disease caused by the pathogen -

XX PS Disclosure; Page 96-97; 125pp; English.

XX CC The invention relates to an immunogenic composition comprising an antigen
 CC derived from a pathogen capable of protecting a host against the
 CC pathogen, mixed with an adjuvant comprising a bleb preparation derived
 CC from a Gram-negative bacterial strain. The immunogenic composition
 CC consists of *N. meningitidis* B blebs or *N. meningitidis* C polysaccharide
 CC antigen. The blebs (derived from the outer membrane) may also
 CC have their toxic lipopolysaccharide (LPS) content reduced using
 CC heterologous down regulating sequences for LPS pathway genes or by
 CC up regulating genes involved in LPS synthesis suppression, by
 CC a promoter replacement technique. The immunogenic preparation is useful
 CC in the manufacture of a medicament for the treatment of a disease caused
 CC by the pathogen from which the antigen is derived (e.g. from
 CC *Neisseria meningitidis* and *bacteraemia*, from *Moraxella*, otitis media and
 CC pneumonia, and from *H. influenzae* chronic bronchitis, sinusitis,
 CC pneumonia and otitis media). The bleb derived from *M. catarrhalis* or


```
DB: 24 Gaps: 0
US-09-912-020-325 (1-477) x ABQ21091 (1-707)
QY 238 SerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyr 257
DB 2 TCCTGTGTAATCGGTAAGCGTCTGTATATGTTAATTTAAGCGTAGGAAGTGAT 61
QY 258 AspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAAGly 277
DB 62 GAGCTANTCGTCGGCGCATACGGGTGATGGCGTTTGGCGTAACGTTGGTAGCGGGT 121
QY 278 AsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyValValGlyLys 297
DB 122 AATTCGTTGGAAGAAGTTGTTTTTTTGTAAATCGCGCGTTGGCGTGGTGGGTAA 181
QY 298 LeuGlyThrSerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAsp 317
DB 182 TTGGGAATTTTACGGTTTCGTCGATCGAGTTGGAATAATGTTTACGTGGACGTGATAG 241
QY 318 ThrGlyPheGlyValMetThrGluGluGluLeuLeuAlaValAlaAlaAATgLy 337
DB 242 ATAGGTTTTCGCGTGATGATCGAAGAGGAATTGAAGTTGCTGATGCGTAGCGGTAA 301
QY 338 ArgGlyGluLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisVal 357
DB 302 CGTGGTGAAAGCGTGATGATTAACGGGTGTTTTTGATATTTGTACGTCGGTAGCTT 361
QY 358 SerThrLeuAlaAsnAlaAlaArgLysLeuGlyAspArgLeuIleValAlaAlaAsnSerAsp 377
DB 362 TTTTATTTGCTAAATGCTCCTAAAGTTGGGTGATCGTTGATTTGCTGTTAATAGCGAT 421
QY 378 AlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMet 397
DB 422 GTTTTATTAACCGTTGAAGGGGATTTTCGTCGGTAAATTTATTCGAATACGGTATG 481
QY 398 IleValLeuGlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrPro 417
DB 482 ATGTGTTGGCGCTATTGGAAGCGGTCGATTTGGTAGTCGTGTTTGAAGAGGATACGTCG 541
QY 418 GlnArgLeuIleAlaGlyIleLeuProAspLeuValLysGlyGlyAspThrLysPro 437
DB 542 TAGCGTTTGTATCGTCGGGATTTTGTAGATTTGTTGGTGAAGCGCGCGATTATAATTA 600
QY 438 GluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsn 457
DB 601 GAAGAGATTCTCGGAGTAAGAAGTTTGGGTTAACGGTGGCGAAGTGTGGTGTAAAT 660
QY 458 PheGluAspGlyCysSerThrThrAsnIleIleLysLysIle 471
DB 661 TTTGAAGACGGTGTTCGACGATTATATATTATTAAGAAGATT 702
RESULT 10
ABQ21091/c
ID ABQ21091 standard; DNA; 707 BP.
XX
AC ABQ21091;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 7682.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
```

```
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
WIPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously,
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 707 BP; 253 A; 230 C; 62 G; 161 T; 1 other;
Alignment Scores:
Pred. No.: 7.37e-70 Length: 707
Score: 856.00 Matches: 184
Percent Similarity: 79.91% Conservative: 3
Best Local Similarity: 78.63% Mismatches: 47
Query Match: 35.77% Indels: 1
DB: 24 Gaps: 0
US-09-912-020-325 (1-477) x ABQ21091 (1-707)
QY 238 SerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyr 257
DB 706 TCCTGTGTTAATCGGTAAGCGTCTGTATATGTTAATTTAAGCGTAGGAAGTGAT 647
QY 258 AspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAAGly 277
DB 646 GACGTANTCGTCGGCGCATACGGTGTGCGCTTTGGCGGTAACTTGGTAGCGGT 587
QY 278 AsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyValValGlyLys 297
DB 586 AATTCGTTGGAAGAAGTTGTTTTTTTGTAAATCGCGCGTTGGCGTGGTGGGTAA 527
QY 298 LeuGlyThrSerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAsp 317
DB 526 TTGGGAATTTTACGGTTTCGTCGATCGAGTTGGAATAATGTTTACGTGGACGTGATAG 467
QY 318 ThrGlyPheGlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAArgLys 337
DB 466 ATAGGTTTTCGCGTGATGATCGAAGAGGAATTGAAGTTGGTGTGGTAGCGGTAA 407
QY 338 ArgGlyGluLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisVal 357
DB 406 CGTGGTGAAAGCGTGATGATTACGGGTGTTTTTCATATTTGTACGTGGGTACGTT 347
```

Qy 358 SerTyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuLeuValAlaValAsnSerAsp 377
 Dbb 346 TTTTATTGGTAAATGTCGTAAGTTGGGTGATCGTTGATGTTGCTTAATAGCAT 287
 Qy 378 AlaSerThrLysArgLeuGlyAspSerArgProValAsnProLeuGluAlaArgMet 397
 Dbb 286 GTTTTATTAAACGGTTCAAGGGGATTTTCGTTCCGTAATTTATTCGAATAGCGTATG 227
 Qy 398 lIeValLeuGlyAlaLeuGluAlaValAspThrProValSerPheGluGluAspThrPro 417
 Dbb 226 ATTGTGTTGGCGGATTCGAAGCGGTGATGGGTAGTGTCTTTGAAGAGGATACGTCG 167
 Qy 418 GluArgLeuAlaGlyLeuProAspLeuValLysGlyAspThrLysPro 437
 Dbb 166 TAGCGTTGATCGCGGATTTGTAGATTTGTGTTGAAGCGGCGGATTAATTA- 108
 Qy 438 GluGluLeuAlaGlySerLysGluValTrpAlaAsnGlyGluValLeuValLeuAsn 457
 Dbb 107 GAAGAGATTGTCGGGAGTAAAGAGTTTGGGTAAACGGTGGCGAAGTGTGTTGTTAAT 48
 Qy 458 PheGluAspGlyCysSerThrThrAsnIleLeuLysLysIle 471
 Dbb 47 TTTGAAGACGGTTGTCACGATTAATATTAATTAAGAAGATT 6
 RESULT 11
 AAA81463/C
 ID AAA81463 standard; DNA: 78845 BP.
 XX AC AAA81463;
 XX DT 04-DEC-2000 (first entry)
 XX DE N. meningitidis partial DNA sequence gnm_11 SEQ ID NO:11.
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 XX KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 XX KW Meningococcus B; MenB; ds.
 XX OS Neisseria meningitidis.
 XX PN WO200022430-A2.
 XX PD 20-APR-2000.
 XX PF 08-OCT-1999; 99WO-US23573.
 XX PR 09-OCT-1998; 98US-0103794.
 XX PR 30-APR-1999; 99US-0132068.
 XX PA (CHIR) CHIRON CORP.
 XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 XX PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 XX PI Rappuoli R, Pizza M;
 XX WPI; 2000-318079/27.
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 XX used in the diagnosis and treatment of N. meningitidis infection and
 XX other Neisserial infections, for example, N.gonorrhoea -
 XX Claim 7; Page 330-353; 1760pp; English.
 XX The present invention describes methods of obtaining immunogenic
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 XX represent specifically claimed Neisseria meningitidis genomic DNA
 XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 XX sequences, which are all used in the exemplification of the present

CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX

SQ Sequence 78845 BP; 19372 A; 21111 C; 19042 G; 19318 T; 2 other;

Alignment Scores:
 Pred. No.: 9,53e-58 Length: 78845
 Score: 753.50 Matches: 159
 Percent Similarity: 67.31% Conservative: 49
 Best Local Similarity: 51.46% Mismatches: 100
 Query Match: 31.49% Indels: 1
 DB: Gaps: 1

US-09-912-020-325 (1-477) x AAA81463 (1-78845)

Qy 8 PheGluArgAlaGlyValMetValValGlyAspValMetLeuAspArgTyrTrpTyrGly 27
 Dbb 46685 TTCGCGCAAGCCAAAGTCTGTGTCGCGACGTGATCGACCGCTATTGGTTCGCG 46626
 Qy 28 ProThrSerArgIleSerProGluAlaProValProValLysValAsnThrIleGlu 47
 Dbb 46625 GATGTGTCCTCGTATTTCGCCCAAGCCCGTGGCGGCAAAATCGGACGAATCGAC 46566
 Qy 48 GluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAla 67
 Dbb 46565 CAACCGCGGGGAGCGGCAAAATGTCGCGGCAACATCGTCTGTTGGCGGAGGCA 46506
 Qy 68 ArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSerLeuAla 87
 Dbb 46505 GGGCTGTTGTCGTAACCGCAACGACGACGCGCGCTCGATCGGTGATGGTG 46446
 Qy 88 AspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLysLeu 107
 Dbb 46445 CAGGACGGCGTCCCTCTATCTGATGCGGCAACAAATCGCCACCGCTCAAACTG 46386
 Qy 108 ArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPheGluGly 127
 Dbb 46385 CGCGTCTGCGCGCAACGACGAGCTTATCGCTCTGATTTGAAGACATCCCAACTGC 46326
 Qy 128 ValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeu 147
 Dbb 46325 GAAAGTGTGGAAACAATCAAGCAGAAATACCGCGAAATCTTGCCGGAATACGACGCAATC 46266
 Qy 148 ValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAla 167
 Dbb 46265 ATTTTTCAGACTACGCAAAAGCGGCTCTGCGCATATCTCCGATATGATGTTGGCGG 46206
 Qy 168 ArgLysAlaGlyValProValIleAspProLysGlyThrAspPheGluArgTyrArg 187
 Dbb 46205 AACACGCGCGCAAAACCGTCTTAATCGCCCAAGCGGACGATACGAAAAATATGTC 46146
 Qy 188 GlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLys 207
 Dbb 46145 GGTGCACTCTGATTACGCTTAACCGCGCAATTTGAAGAAGATGCTGCGGAGTTGGA 46086
 Qy 208 ThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAla 227
 Dbb 46085 AACGAACGAGCTGACCGCAAAAGCGCAAAACCTGCGCGCCCTGACCTGACCTGACCGCC 46026

QY 228 LeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeu 247
 Db 46025 GTTTTACTGACCCGAAGCAAGCATGACCTTGTTCAGGCAAGCGAA---CCGATT 45969
 QY 248 HisMetProThrGlnAlaGlnGluValThrArgValThrGlyAlaGlyAspThrValle 267
 Db 45968 TACCAGCCACCCGCGCCCAAGAGTTTACGACGATATCCGGTGGCGGACACCGTCATT 45909
 QY 268 GlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPheAla 287
 Db 45908 GCCGGAATGGGCTTGGGTCGCGCAGCTGCACCATGCCGCAAGCCATGTACCTTGCC 45849
 QY 288 AsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerProLeuGlu 307
 Db 45848 AATACTGCGCGCGGCTTGTCTGCGCAAACTCGTACGCGGTTGTCTGTTGCCGAA 45789
 QY 308 LeuGluAsnAlaValArgGlyAla 316
 Db 45788 TTGATCAAGGCATTGTTCAGGGCAATCA 45762

RESULT 12
 AAF21608
 ID AAF21608 standard; DNA; 349980 BP.
 XX AC AAF21608;
 XX DT 13-MAR-2001 (first entry)
 XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX OS Neisseria meningitidis.
 XX PN WO200066791-A1.
 XX PD 09-NOV-2000.
 XX PF 08-MAR-2000; 2000WO-US05928.
 XX PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX WPI; 2000-647603/62.
 XX Neisseria meningitidis B full length genome sequence and open reading
 frames are used to detect, treat and prevent Neisserial infections -
 Claim 7; Appendix A; 692pp; English.

The present invention describes the full length genome of
 Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 to AAF21613 represent fragments of the NMB genomic sequence, as the
 sequence was too long to go in a record on its own it was split into 8
 sequences which overlap each other at the beginning and end of each
 sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 AAF21606 represent PCR primers which are used in the exemplification of
 the present invention. The NMB genome and fragments from it have
 antibacterial activity, and can be used in vaccines and gene therapy.
 Neisseria nucleic acids, proteins and/or antibodies which binds to the

CC proteins can be used in compositions for treating or preventing infection
 due to Neisserial bacteria or as a diagnostic reagent for detecting the
 presence of Neisserial bacteria or of antibodies raised to Neisserial
 bacteria. Computers, computer memory, computer storage medium or computer
 databases can be used in a search to identify open reading frames (ORFs)
 or coding sequences within the NMB genome. The DNA sequences provide
 further opportunities to find antigenic or immunogenic proteins which are
 more effective in vaccines than the outer membrane proteins currently
 used.
 XX
 SQ Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

Alignment Scores:
 Pred. No.: 6,58e-57 Length: 349980
 Score: 753.50 Matches: 159
 Percent Similarity: 67.31% Conservative: 49
 Best Local Similarity: 51.46% Mismatches: 100
 Query Match: 31.49% Indels: 1
 DB: 21 Gaps: 1

US-09-912-020-325 (1-477) x AAF21608 (1-349980)

QY 8 PheGluArgAlaGlyValMetValValGlyAspValMetLeuAspArgTyrTrpTyrGly 27
 Db 247885 TTTCGCGCAAGCCAAAGTCTCTGTTGTCGCGACGTGATCTCGACGCTATTGGTTCGCG 247944
 QY 28 ProThrSerArgIleSerProGluAlaProValProValValLysValAsnThrIleGlu 47
 Db 247945 GATGTGTCCTGTTTTCGCCCGAAGCCCGTGGCGGTGGCGAAATCGAGCAATCGAC 248004
 QY 48 GluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAla 67
 Db 248005 CAACGCGCGGCGAGCGCAAAATGTCGCGCGCAACATCGTTCGTTGGCGCGCAGGCA 248064
 QY 68 ArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerIleGlyAla 87
 Db 248065 GGGCTGTTGTCGCTAACCGCAACGACGAGCGCGCGCTCGATCGCTGTTGGTGC 248124
 QY 88 AspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLysLeu 107
 Db 248125 CAGGACGGCGTCCCTCTATCTGATGCGCGCAACAACATCGCCACACCGCTCAACATG 248184
 QY 108 ArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyPheGluGly 127
 Db 248185 CGCGTCGTCGCCCGCAACCGACGCTTATCCGCTCTGATTTTGAAGAATCCCAACTGC 248244
 QY 128 ValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeu 147
 Db 248245 GAAGTGTTCGAAACAATCAAGCAGAAATACCGCGAAATCTTCCCGCAATACGACGCAATC 248304
 QY 148 ValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAla 167
 Db 248305 ATTTTTCAGACTACGCGCAAGCGCGCTGTCGCTATATCCGATATGATGATGCGGCG 248364
 QY 168 ArgLysAlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArgTyrArg 187
 Db 248365 AAACGCGCGCGCAAAACCGCTCTTAATCGACCCCAAGGCGACGATTACCAAAATATGTC 248424
 QY 188 GlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLys 207
 Db 248425 GGTGCAACTCTGATTACGCTAACCGCGCGCAATTCGAAAGAGAGTGTGCGCAGTTGAA 248484
 QY 208 ThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAla 227
 Db 248485 AACGAAGCGAGCTGACCGCAAAAGCGCAAAACCTGCGCGCCACCTCGACCTGACCGCC 248544
 QY 228 LeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeu 247
 Db 248545 GTTTTACTACCGCGAAGCGAAGCATGACCTTGTTCAGCGAAGCGCAA---CCGATT 248601
 QY 248 HisMetProThrGlnAlaGlnGluValThrAspValThrGlyAlaGlyAspThrValle 267
 Db 248602 TACCAGCCACCCGCGCCCAAGAGTTTACGACGATATCGGTGCGGCGCACACCGCTATT 248661

QY 268 GlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhePheAla 287
 Db 248662 GCCGGAATGGCTGGTGGTGGCGGAGGCTGCAACATGCCGAGCCATGTACCTGGC 248721
 QY 288 AsnAlaAlaAlaGlyValValAlaGlyLysLeuGlyThrSerThrValSerProileGlu 307
 Db 248722 AATACTGCGCGCGGGTGTGCTGGCGAAACTCGGTACGCGGTTTCTGCTTTCGCGAA 248781
 QY 308 LeuGluAsnAlaValArgGlyArgAla 316
 Db 248782 TTGATCAAGCATTTGTCAGGCAATCA 248808
 RESULT 13
 ID AAS97221 standard; DNA; 969 BP.
 XX AC AAS97221;
 XX 12-MAR-2002 (first entry)
 XX Neisseria meningitidis virulence gene #26.
 DE Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 KW Infection; Gram-negative bacteria; antimicrobial; ds.
 XX OS Neisseria meningitidis.
 XX PN WC200185772-A2.
 XX PD 15-NOV-2001.
 XX PF 08-MAY-2001; 2001WO-GB02003.
 XX PR 08-MAY-2000; 2000GB-0011108.
 XX PA (MICR-) MICROSCIENCE LTD.
 XX PI Tang C;
 XX WPI: 2002-066593/09.
 XX P-PSDB; ANU72936.
 PT New peptide encoded by operon including virulence genes of *Neisseria meningitidis*, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug -
 PS Claim 1; Page 110-112; 423pp; English.
 XX The invention relates to a peptide (I) encoded by an operon (II) of *Neisseria meningitidis* including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by *Neisseria* or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAS97196-AAS97305 represent *N. meningitidis* virulence genes and related PCR primers of the invention.
 XX SQ Sequence 969 BP; 243 A; 287 C; 262 G; 177 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.74e-59 Length: 969
 Score: 745.50 Matches: 156
 Percent Similarity: 66.34% Conservative: 49
 Best Local Similarity: 50.49% Mismatches: 103
 Query Match: 31.15% Indels: 1
 DB: 24 Gaps: 1
 US-09-912-020-325 (1-477) x AAS97221 (1-969)

QY 8 PheGluArgAlaGlyValMetValGlyValMetLeuAspArgTyrTyrTyrGly 27
 Db 40 TTCGCGCAAGCAAGTCTCTGGTGTGCGGAGCTGATGCTCAGCCGCTATTGGTTCCGC 99
 QY 28 ProThrSerArgIleSerProGluAlaProValProValValLysValAsnThrIleGlu 47
 Db 100 GATGTGTCCCGTATTTCGCCGAAGCCCGTCCGCGTGGCGAAATCGCAATCGAC 159
 QY 48 GluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAla 67
 Db 160 CAACGCGCGGCGGAGCGCAATATGCGCGCAACATCTCGCTGGCGGCAAGTA 219
 QY 68 ArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSerLeuAla 87
 Db 220 GGCCTGTGTGCGGTAAACGGTAACGAGGAGCGCGCTCGACCGCTGATGGTG 279
 QY 88 AspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLysLeu 107
 Db 280 CAGGACGCGCTCGCTCTATCTGATGCGGCAACAAATCGCCACCGCTCAAACTG 339
 QY 108 ArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyPheGluGly 127
 Db 340 CGCGTGTGCGCGCAACAGCAGCTATTCGCTGATTTGAAGAACATCCCAACCGC 399
 QY 128 ValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerIleGlyAlaLeu 147
 Db 400 GAAGTGTGGCAACAATCAAGCGGAATACCGGAAATCTTGCCGCAATACGACGCAATC 459
 QY 148 ValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAla 167
 Db 460 ATTTTTCAGACTACGCAAGGCGCTGTGCGACATCTCCGATATGATGATGGCG 519
 QY 168 ArgLysAlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArgTyrArg 187
 Db 520 AAACAGGAGCAAAACCGTATTATCGACCCCAAGCGCAGCTATCGCAAAATAGCC 579
 QY 188 GlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValAlaGlyLysCysLys 207
 Db 580 GCGCAACCTGATTACGCGCAACCGCGGAATTTGAAGAAGTGTGCGCAGCTGAAA 639
 QY 208 ThrGluGluLeuValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAla 227
 Db 640 AAGCAAAATGATTGACCGAAAGCCCAAAACCTGCGCGCCACCTCGACTTGACCGCT 699
 QY 228 LeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeu 247
 Db 700 ATTTTACTGACCGAGCGAAGAGGATGACCTTTGTTACGGAAGCGCAA--CCCAT 756
 QY 248 HisMetProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValle 267
 Db 757 TACGAGCCACCGCGCGCAAGAGTTTACGACGTATCCGCGTGGCGGCGACACCGCTCAT 816
 QY 268 GlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhePheAla 287
 Db 817 GCGGGAATGGCTGGTGGCGGAGGCTGACCATGCCGCAAGCCATGATGATCTTGGC 876
 QY 288 AsnAlaAlaAlaGlyValValAlaGlyLysLeuGlyThrSerThrValSerProileGlu 307
 Db 877 AATACTGCGCGCGGGTGTGCTGGCGAAACTCGGTACGCGGTTTCTGCTTTCGCGAA 936
 QY 308 LeuGluAsnAlaValArgGlyArgAla 316
 Db 937 TTGACCAAGCATTTGTCAGGCAATCA 963
 RESULT 14
 ID AAT68016
 XX AAT68016 standard; DNA; 1272 BP.
 AC AAT68016;
 XX 15-JUL-1997 (first entry)
 XX H. pylori cytoplasmic protein ORF 07cell1409orf4.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 XX Helicobacter pylori.

OS
 FH Key Location/Qualifiers
 FT CDS 1..1272
 FT /*tag= a
 FT /note= "no stop codon given in sequence"

XX W09640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglinth OT, Smith D, Melligaard BL;

XX WPI; 1997-052306/05.

XX P-PSDB; AAW20763.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

XX Claim 9; Pages 847; 1481pp; English.

XX The present sequence encodes a Helicobacter pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX Sequence 1272 BP; 412 A; 232 C; 284 G; 344 T; 0 other;

XX Alignment Scores:

Pred. No.: 2.56e-58 Length: 1272
 Score: 734.50 Matches: 177
 Percent Similarity: 58.66% Conservative: 77
 Best Local Similarity: 40.88% Mismatches: 154
 Query Match: 30.69% Indels: 25
 DB: 18 Gaps: 9

US-09-912-020-325 (1-477) x AAT68016 (1-1272)

Qy 51 GlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeuVal 70
 Db 7 GCGGAGCGGCATATGGCTTAACCTTACCTCTTAAAGCTAAAGCTTTTATGT 66
 Qy 71 GlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSerLeuAlaAspValAsn 90
 Db 67 GGGGTAGTGGGAGATGATTTAAAGGCAAGCATTTTACCTTTAAATCTATAAGG 126
 Qy 91 ValLysCysAspPheValSerVal---ProThrHisProThrIleThrLysLeuArgVal 109
 Db 127 ATTGACACTTCAGGTGTTTAAATAGATAAAACCCGTTGCACCGCTTAAACGCGCATC 186

Qy 110 LeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPheGluGlyValAsp 129
 Db 187 ATCGCGCAAAACCAACGAAATCGTCGCGTGGATAAGGAATC-----AAAGAC 234
 Qy 130 Pro-----GlnProLeuHisGluArgIleAsnGlnAlaLeuSerSer 143
 Db 235 CCCTTAAACGCTGATTTAAGAAAGAAATCTTTTAGATTTTATCGCAGAAATAATCAAGAA 294
 Qy 144 IleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeu-----AlaSerValGln 161
 Db 295 ATAGATGCGGATCCCTTTTACAGATTACAATGAAGGTGTGTGGATTTTGAACCTCACTCAA 354
 Qy 162 GlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyThr 181
 Db 355 ACCATCATCGCTAGCTAATAAGCATCATAGCTCATTTTATCGCAACCTTAAAGGAAG 414
 Qy 182 AspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAla 201
 Db 415 GATTATAGCAAAATATCCCATCGAGTTTGTATCAAGCTTAATCGCTGAATTAGAGCAA 474
 Qy 202 ValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIleAla 221
 Db 475 GCGCTCCATTTGAAATTAGACAGCCATCGGAATTTATCAAAAGCGCTCCAATTTTACAA 534
 Qy 222 Asp---TyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
 Db 535 GAAACTTATCATATCGCTATGCCCTTAGTAACCTTGAGCGAACAAGCATCGCTTTTAA 594
 Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
 Db 595 GAAAAAGCGAGTTA--GTCAATTGCCCACTATCGCTAAAGAAAGTTTATGATGTAACG 651
 Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
 Db 652 GGGGAGCGGATACGGGTATAGCTCTTTTAAACGCTCTCTTTTGGAAATCAAAAAGCTTG 711
 Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
 Db 712 AAAGAAGCTTCGAGTTTGTCTAATCGCGCTCGGGGTGGTGGTGGTAAATGGGAGC 771
 Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
 Db 772 GCGTTAGCGAGT-----TTAGAGAAATCGCTTTTGAACCAACGACGCCCT 822
 Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
 Db 823 AAAATCCTCCCTTTAGAAAAGCTG-----TTAGAACTTTAGAACGCAACCAAGCAA 873
 Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
 Db 874 AAAATCGTTTTCACCAATGGCTGTTTTCATATTCTCCATAAAGGCGATGCGAGCTATTG 933
 Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
 Db 934 CAAAAGGCTAAAGCTTTTAGGGGATATTCTTGTGTGGGGTTTAAATAGCGATAAATTCAT 993
 Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
 Db 994 AAAAGGCTTAAGGGGATAAAGCCCTTAGTGAGGAAAAAGACAGCGCGCTTCTTTTA 1053
 Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
 Db 1054 GCGAGCTTGCTTCGCTGGATTATGTTGTGTGTTGGAGAAGACACGCCCAATAAAATG 1113
 Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
 Db 1114 ATTCAAGCCCTAAAGCCCTGATATTATTAGTCAAGGAGCGGACTACCTCAATAAAGAAAGTC 1173
 Qy 441 AlaGlySerLysGluValTyrAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 Db 1174 ATAGGGAGCGAG-----TTGGCTAAAGAACCCCGTTTGTATAGAAATTTGAAGAA 1221
 Qy 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGln 473

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 26, 2002, 10:25:42 ; Search time 2404 Seconds
(without alignments)
3213.497 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERGVVGVDM.....FEDGCTTIIKKIQDKKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPRO_pool/US09912020/runat_25112002_091428_24259/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09912020.qcgn.1.1.2024 -runat_25112002_091428_24259 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST :
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_oth.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	C	290	12.1	505	10	BE332786	BE332786 us53d03.x
2	C	233	9.7	483	17	BH404405	BH404405 AG-ND-140
3	C	220	9.2	550	17	BH390803	BH390803 AG-ND-140
4		179	7.5	502	13	BM167110	BM167110 EST569633
5		178.5	7.5	555	13	BT786239	BT786239 sai34e10.
6		178	7.4	652	14	BQ996249	BQ996249 QGG12F01.
7		178	7.4	1031	11	AK017229	AK017229 Mus muscu
8		173.5	7.3	572	13	BI943547	BI943547 sq63e05.y
9		172.5	7.2	736	14	BQ850159	BQ850159 QGB11N07.
10		171.5	7.2	629	13	BM169900	BM169900 EST572423
11	C	167.5	7.0	689	13	BJ165221	BJ165221 BJ165221
12		166.5	7.0	485	13	BT944705	BT944705 sad41q05.
13		162	6.8	686	14	BQ995975	BQ995975 QGG11110.
14		158	6.6	729	13	BI308004	BI308004 EST529414
15		157	6.6	609	12	BF642033	BF642033 NF014A091
16		156	6.5	603	12	BG140383	BG140383 EST480825
17	C	156	6.5	631	13	BJ435742	BJ435742 BJ435742
18		156	6.5	668	13	BJ428649	BJ428649 BJ428649
19	C	156	6.5	682	13	BJ432922	BJ432922 BJ432922
20	C	156	6.5	684	13	BJ432289	BJ432289 BJ432289
21		156	6.5	685	14	C92444	C92444 C92444 Dict
22		156	6.5	695	12	BF154250	BF154250 057F11 Ma
23		156	6.5	699	13	BJ432465	BJ432465 BJ432465
24		156	6.5	705	13	BJ377283	BJ377283 BJ377283
25		156	6.5	745	13	BJ431261	BJ431261 BJ431261
26		156	6.5	750	13	BJ372557	BJ372557 BJ372557
27	C	156	6.5	784	13	BJ428867	BJ428867 BJ428867
28	C	155.5	6.5	568	9	AI995602	AI995602 701675044
29	C	155	6.5	649	13	BJ435218	BJ435218 BJ435218
30		154.5	6.5	603	10	AV828468	AV828468 AV828468
31		154.5	6.5	792	12	BG888833	BG888833 EST514684
32		154	6.4	565	9	AU262555	AU262555 AU262555
33		154	6.4	768	12	BG131085	BG131085 EST463977
34		152.5	6.4	523	13	BM093385	BM093385 sai08d02.
35		152.5	6.4	537	12	BE920533	BE920533 EST424302
36	C	152.5	6.4	698	12	BG145720	BG145720 uu90b08.x
37		152.5	6.4	707	12	BG440694	BG440694 GA_Ea000
38		152	6.4	648	13	BM358455	BM358455 GA_Ea000
39	C	151.5	6.3	695	14	BQ198857	BQ198857 UI-R-ES0-
40		151	6.3	507	13	BI427212	BI427212 sah77f06.
41	C	150.5	6.3	528	9	AI326806	AI326806 mjl6e06.x
42		150.5	6.3	684	12	BF453934	BF453934 maa60g10.
43	C	149.5	6.2	553	10	AW219042	AW219042 EST301524
44		149.5	6.2	616	13	BT639334	BT639334 SD1893.5
45	C	148.5	6.2	528	10	BE100952	BE100952 UI-R-BJ1-

ALIGNMENTS

RESULT 1
BE332786
LOCUS us53d03.x1 Perkins LRH Mus musculus cDNA clone IMAGE:3216101 3',
DEFINITION similar to SW:RFAE_ECOLI P76658 ADP-HEPTOST: SYNTHASE ;, mRNA
sequence.
ACCESSION BE332786
VERSION BE332786
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 505)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

QGG12F01, mRNA sequence.
 BO996249
 BO996249.1 GI:22430645
 EST.
 Lactuca sativa.
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.
 1 (bases 1 to 652)
 Kozik,A., Micheltore, R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Micheltore]
 Department of Vegetable Crops, R.W.Micheltore Lab
 University of California at Davis (UCD)
 Asmundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [micheltore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig5995, see http://cgpdb.ucdavis.edu/
 for details.
 Plate: QGG12 row: F column: 01.
 Location/Qualifiers
 1. 652
 /organism="Lactuca sativa"
 /cultivar="L. serriola"
 /db_xref="taxon:4236"
 /clone="QG12F01"
 /lab_host="E.coli"
 /note="Vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_L18-QS_EFGHJ lettuce serriola
 TAG_TISSUE=flowers pre-fertilized
 TAG_SEQ=GGTTGACGGG"
 BASE COUNT 182 a 123 c 179 g 168 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.67e-10 Length: 652
 Score: 178.00 Matches: 64
 Percent Similarity: 43.13% Conservative: 27
 Best Local Similarity: 30.33% Mismatches: 60
 Query Match: 7.44% Indels: 60
 DB: 14 Gaps: 8
 US-09-912-020-325 (1-477) x BO996249 (1-652)
 Qy 297 LysLeuGlyThrSerThrValSerProIleGluLeuGluAsnAlaValArg----- 313
 Db 39 ANGATGGGTTCGGATACA-----GGTCTAGAGAACTCGAAGCGGTTCACGC 86
 Qy 314 -----GlyArgAlaAspThrGlyPheGlyValMet-----Thr 324
 Db 87 ACATGGTTGATCGGAGGTTTGGTGCAGCGCTATCGGTTATGGGTTATCGCTCTGG 146
 Qy 325 GluGluGluLeuLysLeuAlaAlaAlaArgLysArgGlyGluLysValValMet 344
 Db 147 CCGGAGCTCTCCCTCTGAGTATTGTTGGGAGGAGAGAGAGAGCGCCCATTCGTGTATAC 206

Qy 345 ThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaArg 364
 Db 207 ATGGATGCTGCTTTGATATGTCATATATGTCACACGCACTCGACACACGACGT 266
 Qy 365 LysLeuGlyAspArgLeuLeuValAlaValAsnSerAspAlaSerThrLysArgLeuLys 384
 Db 267 GCTCTCGGAGATCAATGATTGTTGGCGTTGTCACGCGATGCTGAAATCATTCGAACAAA 326
 Qy 385 GlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeuGluAlaLeuGlu 404
 Db 327 GGCCCT-----CCGGTTACCCCTCTTCATGAAGAGATGATTGTTGAGTCTGTGAAA 380
 Qy 405 AlaValAspTrpValValSer----- 411
 Db 381 TGGGTGGATGAAGTCATTCACAGATGCACCATATGCCATATGCAACTGAGAAATTCATGAGGAAG 440
 Qy 412 ---PheGluGluAspThrProGlnArgLeuIleAlaGly-----IleLeuPro 426
 Db 441 CTTTTTGTGATGATACACATAGATTACATTCATGATGAGATGATCCATGCAATTCCTCCT 500
 Qy 427 Asp-----LeuLeuValLysGlyGlyAspTyrLysProGluGluIle 440
 Db 501 GATGAAGTGCAGCATATGCCCCCTTCGCAAGAAAGCAGGTTCGATATAGCAA----- 551
 Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGluValLeuValLeuAsnPheGluAsp 460
 Db 552 -----ATCAACGACACAGAA 566
 Qy 461 GlyCysSerThrThrAsnIleIleLysIle 471
 Db 567 GGAGTCTCAAGCACACAGCATTTGTTGGGAGGATG 599

RESULT 7
 AKO17229 1031 bp mRNA linear HTC 19-JAN-2002
 LOCUS Mus musculus adult male xiphoid cartilage cDNA, RIKEN full-length
 DEFINITION enriched library, clone:5230400M11:homolog to RIBOKINASE, full
 insert sequence.
 ACCESSION AKO17229.1 GI:12856364
 VERSION AKO17229.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J).adult male xiphoid cartilage cDNA to
 mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
 clone:5230400M11.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Cerninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 2
 Cerninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042155
 3
 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Chera,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multipillar sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913

QY 248 sMetProThrGlnAlaGlnGluValTyAspValThrGlyAlaGlyAspThrValIleG1 268
 Db 790 CATTCCCACTGAAGCTGTCAAGCTGTAGACACACCGGCTGTGTGACAGTTTCTGGG 849
 QY 268 yValLeuAlaAlaThrLeuAla-----AlaGlyAsnSerLeuGluGluAlaCysPhePh 286
 Db 850 AGCGTGGCGCTTCTACCTGGCTTACTACCAAACTGTCTCTTGAAGAATACTCTCAAGAG 909
 QY 286 eAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
 Db 910 ATCCAAATTCATCGTGCAGTCAGTGTACAGCCACGGGAAACACAGCTCTTATCCCA 967

RESULT 8
 B1943547

LOCUS B1943547 572 bp mRNA linear EST 06-DEC-2001
 DEFINITION sq3e05.y1 Gm-cl048 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl048-153 5' similar to TR:Q9SHH5 Q9SHH5 F20D23.14 PROTEIN. ;
 mRNA sequence.

ACCESSION B1943547
 VERSION B1943547.1 GI:16279653
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max
 soybean.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 572)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
 A., Bolla,B., Marr,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Willson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 409.
 Location/Qualifiers
 1..572
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl048-153"
 /tissue_type="whole seedlings of greenhouse grown plants"
 /dev_stage="1 week old"
 /lab_host="DH10B"
 /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; The Clark NIL was constructed and seed was provided
 by Dr. J. Specht, University of Nebraska (Shoemaker and
 Specht, 1995). The cDNA library was constructed from mRNA
 isolated from whole seedlings of 1 week old greenhouse
 grown plants. Complementary DNA was synthesized from mRNA
 using a primer consisting of a poly(dT) sequence with a
 XhoI restriction site and a 3' anchor. EcoRI adapters were
 ligated to the blunt-ended cDNA fragments followed by XhoI
 digestion. The cDNA fragments were directionally cloned
 into the EcoRI-XhoI restriction site of the pBluescript
 vector. The ligated cDNA fragments were transformed into
 DH10B host cells (GibcoBRL). The library was constructed
 in cooperation with Dr. Paul Keim's laboratory at Northern
 Arizona University."

BASE COUNT 134 a 124 c 153 g 159 t 2 others

ORIGIN

Alignment Scores: 1.26e-09 Length: 572
 Pred. No.: 173.50 Matches: 53
 Score: 48.03% Conservative: 20
 Percent Similarity: 34.87% Mismatches: 68
 Best Local Similarity: 7.25% Indels: 11
 Query Match: 13 Gaps: 5
 DB: 13

US-09-912-020-325 (1-477) x B1943547 (1-572)

QY 162 GlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuLeuLeuLeuLeuLeuLeu 181
 Db 99 CAAGTCGCACAGCTGCAAGGAATGCTGCGCTGCCAGTAGTGTGTGATCGCGGGGCATG 158
 QY 182 AspPhe-----GluArgTyArgGlyAlaThrLeuLeuThrProAsnLeuSer 197
 Db 159 GATGGGCCCTTCCACCAATATTATTGATTTGTTGATATTCTGAGTCTTAATGAAACT 218
 QY 198 GluPheGluAlaValGly---LysCysLysThrGluGluGluIleValGluArgGly 216
 Db 219 GAATTCGTGCGCTTACCGGAATGCCAAGAAAGTTTGAAGAGATTCCACAGCGTCT 278
 QY 217 MetLysLeuIleAlaAspTyrGluLeuSerAla-----LeuLeuValThrArgSerGlu 234
 Db 279 TTGAATGC-----CATGAATGGGAGTTAAGCAAGTCTTGTGAAACTTGGGAA 329
 QY 235 GlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThr---GlnAla 253
 Db 330 AAAGATCTGCGCTTTTGTAGAAGGAGAGAAACCAATTCAGCAGCTGCCATCTGCT 389
 QY 254 GlnGluValTyAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThr 273
 Db 390 AAACAGTCGTGTGATACTACTGTCGCGGTGATACCTTTTACTGCTGCTTTTGTGTGGCC 449
 QY 274 LeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyVal 293
 Db 450 TTGGTTGAGGGCAAGTCCANAAAGGAATCCCTCAGATNTGCTGCTCGGCTGTCTCT 509
 QY 294 ValValGlyLysLeuGlyThrSerThrValSerPro 305
 Db 510 TGTGTTCAATGAAGGAGGCGCTCTCTCTAGCATGCT 545

RESULT 9
 BQ850159

LOCUS BQ850159 736 bp mRNA linear EST 14-AUG-2002
 DEFINITION QGB11N07.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 QGB11N07, mRNA sequence.

ACCESSION BQ850159
 VERSION BQ850159.1 GI:22235628
 KEYWORDS EST
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa.

REFERENCE 1 (bases 1 to 736)
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevaller,F., Ziegler,J., Ellison
 P., Koikman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lactucae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactucae.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: a.kozik@atgc.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QG_CA_Contig5995, see <http://cgpdb.ucdavis.edu/> for details.
Plate: QGB11 row: N column: 07.

FEATURES

source
1. 736
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGB11N07"
/lab_host="E. coli"
/note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG_LIB=QG_ABCDI lettuce salinas
TAG_TISSUE=germinating seeds
TAG_SEQ=TCTGTGCGG"

BASE COUNT 195 a 147 c 197 g 197 t
ORIGIN

Alignment Scores:
Pred. No.: 2,49e-09 Length: 736
Score: 172.50 Matches: 62
Percent Similarity: 45.41% Conservative: 27
Best Local Similarity: 31.63% Mismatches: 59
Query Match: 7.21% Indels: 48
DB: 14 Gaps: 7

US-09-912-020-325 (1-477) x BQ850159 (1-736)

Qy 303 ValSerProfilLeuLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyVal 322
Db 13 GTTTCTAGCCACATGGTTGATGGAGGCTTTGGTGGCAGGGCT---ATCGGTTATGGGTT 69
Qy 323 Met-Thr-----GluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGly 340
Db 70 ATGCGCTGTTGGCGGAGCTCTCCCTCGATGATTGTTGGGAGGAAGAAGAGGCC 129
Qy 340 uLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyLe 360
Db 130 CATTCGTGTATACATGATGCTCTTGATATGATGATGATGATGATGATGATGATGAT 189
Qy 360 uAlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerTh 380
Db 190 CCGACAAGCAGCTGCTCGGAGATCAATTGATTGTTGGGCTTGTGCGATGCTGAAT 249
Qy 380 rLysArgLeuLysGlyAspSerArgProValAspProLeuCluclnArgMetIleValLe 400
Db 250 CATTGCAAAACAAGGCCCT-----CGGTATGCCCTCTTCATGAAGAGGATGATTGGT 303
Qy 400 uGlyAlaLeuGluAlaValAspTrpValValSer----- 411
Db 304 GAGTGCTGTGAATGGGTAGATGAAGTCATTCAGATGCCATATGCCATTAACATGAAGA 363
Qy 412 -----PheGluGluAspThrProGlnArgLeuIleAlaGly----- 423
Db 364 ATTCATGAGGAAGCTTTTGTATGAATACAACATGATATCATTCATGAGATGATGCC 423
Qy 424 ----IleLeuProAsp-----LeuLeuValLysGlyGlyAspTrpLy 436
Db 424 ATGCATTCTCCCTGATGGAAGTACGCGATATGCCCTTGCAAAAGAAAGCAGGTCGATATA 483
Qy 436 sProGluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLe 456
Db 484 GCAA----- 489
Qy 456 uAsnPheCluAspGlyCysSerThrThrAsnIleIleLysLysIle 471

Db 490 CAACAGAACAGAGGAGTCTCAAGCACAGACATTTGTTGGGAGGATG 535

RESULT 10

BM169900

LOCUS

DEFINITION

BM169900

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Plasmodium yoelii yoelii

Plasmodium yoelii yoelii

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

1 (bases 1 to 629)

Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,

Fraser,C.M. and Carucci,D.J.

Plasmodium yoelii EST project at TIGR

Unpublished (2001)

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

<http://www.malaria.mr4.org/mr4pages/index.html>

Seq primer: ADF.

Location/Qualifiers

1..629

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCPY36"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with PY17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybrizAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybrizAP vector and plasmid DNA

isolated."

BASE COUNT 268 a 73 c 100 g 188 t

ORIGIN

Alignment Scores:

Pred. No.: 2,52e-09 Length: 629

Score: 171.50 Matches: 51

Percent Similarity: 46.43% Conservative: 40

Best Local Similarity: 26.02% Mismatches: 64

Query Match: 7.17% Indels: 41

DB: 13 Gaps: 7

US-09-912-020-325 (1-477) x BM169900 (1-629)

Qy 297 LysLeuGlyThrSerThrValSerProfilLeuLeuGluAsnAlaValArgGlyArgAla 316

Db 101 AAAAAAAAAACAGTGTATGCCA-----TCAATATATAAAGGTTTAAATG 148

Qy 297 LysLeuGlyThrSerThrValSerProIleGluLeuGluAsnAlaValArg----- 313
 |||:||||:|||||
 Db 76 AAGATGGTTCCGATACA-----CGCTAGAGAACTCGAACGGTTTCCNCG 123
 |||:||||:|||||
 Qy 314 -----GlyArgAlaAspThrGlyPheGlyValMet-----Thr 324
 |||:||||:|||||
 Db 124 ACATGGTTGATCGGAGGTTTGGTGGCAGGCTATCGGTTATGGGTTATGCGTGTGG 183
 |||:||||:|||||
 Qy 325 GluGluGluLeuLysLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaValMet 344
 |||:||||:|||||
 Db 184 CCGAGCTCTCCCTCTGATGTTGTTGGGAGGAAGAAGAGAGCCCATTCGTTGATAC 243
 |||:||||:|||||
 Qy 345 ThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaArg 364
 |||:||||:|||||
 Db 244 ATGATGGCTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 303
 |||:||||:|||||
 Qy 355 LysLeuGlyAspArgLeuLeuValAlaValAsnSerAspAlaSerThrLysArgLeuLys 384
 |||:||||:|||||
 Db 304 GCTCTCGGAGATCAATTGATGTTGGCTGTCAGCGATGCTGAAATCATTCGCAACAA 363
 |||:||||:|||||
 Qy 385 GlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeuGlyAlaLeuGlu 404
 |||:||||:|||||
 Db 364 GCCTCT-----CCGTTACCCCTCTTCATGAAGGATGATGATGATGATGATGATGAT 417
 |||:||||:|||||
 Qy 405 AlaValAspTrpValVal-----SerPheGluGluAspThrProGlnArg 419
 |||:||||:|||||
 Db 418 TGGTAGATGAAGTCATTCAGATGCACCATATGCCATATCCATGAAGATTCATGAGGAAC 477
 |||:||||:|||||
 Qy 420 LeuIleAlaGlyLeuLeuPro-AspLeuLeuValLysGlyAspTyrLysProGluGlu 439
 |||:||||:|||||
 Db 478 GCCCTNNNTCAATACAAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
 |||:||||:|||||
 Qy 439 uIleAlaGlySerLysGluValTrpAla-----AsnGlyGlyGluValLeuValLe 456
 |||:||||:|||||
 Db 532 TCTCCTGATGAATGACGATATGCTGCTGCAAGAAAGACGATGATATGAACAAAT 591
 |||:||||:|||||
 Qy 456 uAsnPheGluAspGlyCysSerThrThrAsnIleLeuLysIle 471
 |||:||||:|||||
 Db 592 CAACAGACAGAGAGCTCAAGCAGACAGATGTTGGGAGGATG 637
 |||:||||:|||||

RESULT 14
 BI308004
 LOCUS EST529414 GP0D Medicago truncatula cDNA clone pGPOD-1H20 5' end,
 DEFINITION mRNA sequence. EST 20-JUL-2001
 ACCSSION BI308004
 VERSION BI308004.1 GI:14982331
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 729)
 AUTHORS Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T., Cho
 J., and Fraser, C.M.
 TITLE ESTs from developing reproductive tissues of Medicago truncatula
 JOURNAL Unpublished (2001)
 COMMENT Contact: Michael A. Grusak
 USDA/ARS Children's Nutrition Research Center
 Baylor College of Medicine
 1100 Bates Street, Houston, TX 77030-2600, USA
 Tel: 713-798-7044
 Fax: 713-798-7078
 Email: mgrusak@bcm.tmc.edu
 B394859e
 TIGR sequence name: MTOAD46TK
 More information is available at: www.medicago.org
 Seq primer: Skmod (CTA gAA CTA gTG gAT CC).
 Location/Qualifiers
 FEATURES
 source 1..729

/organism="Medicago truncatula"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="pGPOD-1H20"
 /clone_lib="GPOD"
 /tissue_type="Immature pod walls"
 /dev_stage="Immature pods, ranging in age from 15 to 30
 days after pollination"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Immature pods, ranging in age from 15 to 30 days
 after pollination, were collected from greenhouse-grown
 plants. At harvest, seeds were removed from pods and
 isolated pod walls were collected and immediately frozen
 in liquid nitrogen. Pod walls were pooled for mRNA
 extraction. cDNA was prepared from polyA+ enriched RNA.
 The cDNA was directionally ligated into the Unizap XR
 vector from Stratagene and packaged using Gigapack III
 Gold packaging extracts. Plasmids containing cDNA inserts
 were excised from the recombinant lambda-zap phage using
 Ex-assist helper phage and propagated in XL0LR cells."
 BASE COUNT 186 a 110 c 190 g 243 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.2e-07 Length: 729
 Score: 158.00 Matches: 46
 Percent Similarity: 52.74% Conservative: 31
 Best Local Similarity: 31.51% Mismatches: 55
 Query Match: 6.60% Indels: 14
 DB: 13 Gaps: 5

US-09-912-020-325 (1-477) x BI308004 (1-729) 1

Qy 336 ArgLysArgGlyGluLysVal-----ValMetThrAsnGlyValPheAspIleLeuHis 353
 |||:||||:|||||
 Db 130 AGAAGAAGGAATAAAACCCGATTCGCTTTATATGGATGGTGTGTTGATATGATGATCAT 189
 |||:||||:|||||
 Qy 354 AlaGlyHisValSerTyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuIleValAla 373
 |||:||||:|||||
 Db 190 TATGCCCATTTGTAATGCTTTCGTCACGCTCGTCTCTGTGGTATCAATGATGTTGGT 249
 |||:||||:|||||
 Qy 374 ValAsnSerAspAlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeu 393
 |||:||||:|||||
 Db 250 GTTGTGTAGTATGATGAATAATTATTCCTAATAAGGCTCCT-----CCTGTTACCCCTTTA 303
 |||:||||:|||||

Qy 394 GluGlnArgMetIleValLeuLeuGluAlaValAspTrpValVal----- 410
 |||:||||:|||||
 Db 304 CATGAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
 |||:||||:|||||
 Qy 411 -----SerPheGluGluAspThrProGlnArgLeuIleAlaGlyIleLeuProAspLeu 428
 |||:||||:|||||
 Db 364 CCATATGCCATTACTGAGGAATTCATGAAGAAGCCTTTTCGATGAGTATAATATAGATTAC 423
 |||:||||:|||||
 Qy 429 LeuValLysGlyGlyAspTyrLysProGluGluIleAlaGlySerLysGluValTrpAla 448
 |||:||||:|||||
 Db 424 ATTATTCATGGGAGCAT-----CCTTGTGTTCTCTGATGGAACCTGATCCTTATGCT 477
 |||:||||:|||||

Qy 449 Asn-----GlyGlyGluValLeuValLeuAsnPheGluAspGlyCysSerThrThr 465
 |||:||||:|||||
 Db 478 CATGCCAAAGAGGCTGGCGCTATTAACAGATTAAGCGACGAGGAGGAGTTCACGACT 537
 |||:||||:|||||

Qy 466 AsnIleIleLysIle 471
 |||:||||:|||||
 Db 538 CATATTGTGGTTCGAATG 555
 |||:||||:|||||
 RESULT 15
 BF642033
 LOCUS BF642033 609 bp mRNA linear EST 19-DEC-2000
 DEFINITION NF014A09IN1F1068 Insect herbivory Medicago truncatula cDNA clone
 NF014A09IN 5', mRNA sequence.
 ACCSSION BF642033
 VERSION BF642033.1 GI:11906191
 KEYWORDS EST.

GenCore version 5.1.3
Copyright (c) 1993 - 2002. Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: November 26, 2002, 10:27:12 : Search time 73 seconds
(without alignments)
2003.902 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFAGERVMVGDV.....FEDGCGTNIKKIQDKKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09912020/runat_25112002_091429_24273/app_query.fasta_1.647
-DB=Issued_Patents_NA -qfmt=fastap -suffix=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09912020.cgn1.1.38 @runat_25112002_091429_24273 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query		Length	DB	ID	Description		
		Match	%						
1	411.5	17.2		347	4	US-08-651-155B-185	Sequence 185, App		
2	263	11.0		294	4	US-08-651-155B-186	Sequence 186, App		
3	165.5	6.9		948	4	US-09-134-001C-1684	Sequence 1684, Ap		
4	151	6.3		3796	1	US-08-920-812-19	Sequence 19, Appl		
5	151	6.3		3796	1	US-08-920-827-19	Sequence 19, Appl		
6	151	6.3		3796	1	US-08-921-177-19	Sequence 19, Appl		
7	151	6.3		3796	1	US-08-362-577C-19	Sequence 19, Appl		
8	151	6.3		3796	2	US-08-920-828-19	Sequence 19, Appl		
9	150	6.3		414	4	US-09-134-001C-2723	Sequence 2723, Ap		
c 10	146.5	6.1		12804	4	US-09-453-702B-149	Sequence 149, App		
c 11	138.5	5.8		2580	4	US-09-199-637A-280	Sequence 280, App		
c 12	138.5	5.8		2970	4	US-09-199-637A-272	Sequence 272, App		

13	138.5	5.8	3058	4	US-09-221-017B-757		Sequence 757, App		
c 14	138.5	5.8	42235	4	US-09-199-637A-1		Sequence 1, Appl		
15	137.5	5.7	939	4	US-09-134-001C-2746		Sequence 2746, Ap		
16	137.5	5.7	1907	3	US-08-826-611-1		Sequence 1, Appl		
17	134.5	5.6	44377	2	US-08-804-227C-7		Sequence 7, Appl		
18	134.5	5.6	44377	2	US-08-804-198-1		Sequence 1, Appl		
19	129.5	5.4	939	4	US-08-487-429A-5		Sequence 43, Appl		
20	128	5.3	1407	3	US-08-714-918-43		Sequence 43, Appl		
21	128	5.3	1407	4	US-09-265-315-43		Sequence 43, Appl		
22	128	5.3	1407	4	US-09-265-315-43		Sequence 43, Appl		
23	128	5.3	1407	4	US-09-266-417-43		Sequence 43, Appl		
c 24	127	5.3	49377	1	US-08-784-233A-1		Sequence 1, Appl		
c 25	125	5.2	28958	1	US-08-258-261B-6		Sequence 6, Appl		
c 26	125	5.2	28958	1	US-08-456-837-6		Sequence 6, Appl		
c 27	125	5.2	28958	1	US-08-457-342-6		Sequence 6, Appl		
c 28	125	5.2	28958	1	US-08-457-646A-6		Sequence 6, Appl		
c 29	125	5.2	28958	1	US-08-458-076A-6		Sequence 6, Appl		
c 30	125	5.2	28958	1	US-08-764-233A-4		Sequence 4, Appl		
c 31	125	5.2	28958	1	US-08-457-335A-6		Sequence 6, Appl		
c 32	125	5.2	28958	1	US-08-729-214-6		Sequence 6, Appl		
c 33	125	5.2	28958	3	US-09-028-934-6		Sequence 6, Appl		
c 34	122.5	5.1	1190	2	US-08-479-614-3		Sequence 3, Appl		
c 35	122	5.1	11471	4	US-09-504-358-16		Sequence 16, Appl		
36	117.5	4.9	1181	2	US-09-954-314-16		Sequence 16, Appl		
c 37	117.5	4.9	1181	2	US-08-479-614-9		Sequence 9, Appl		
39	116.5	4.9	1190	2	US-08-479-614-1		Sequence 1, Appl		
40	116.5	4.9	4848	4	US-08-955-957A-1		Sequence 1, Appl		
41	116.5	4.9	4848	4	US-08-955-957A-4		Sequence 4, Appl		
42	116.5	4.9	4848	4	US-08-955-957A-6		Sequence 6, Appl		
43	116	4.8	1279	3	US-08-826-611-5		Sequence 5, Appl		
c 44	116	4.8	1510	1	US-08-361-920-26		Sequence 26, Appl		
c 45	116	4.8	1510	1	US-08-479-939-26		Sequence 26, Appl		

ALIGNMENTS

RESULT 1

US-08-651-155B-185
Sequence 185, Application US/08651155B
Patent No. 6365401

GENERAL INFORMATION:
APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiechoff Mr., Douglas M.
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
INFECTION
TITLE OF INVENTION: INFECTION
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Chrisman, Bynum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651.155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen Mr., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475

; INFORMATION FOR SEQ ID NO: 185:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-651-155B-185

Alignment Scores:

Pred. No.: 6,1e-37 Length: 347
Score: 411.50 Matches: 99
Percent Similarity: 88.14% Conservative: 5
Best Local Similarity: 83.90% Mismatches: 8
Query Match: 17.20% Indels: 7
DB: 4 Gaps: 2

US-09-912-020-325 (1-477) x US-08-651-155B-185 (1-347)

QY 22 AspArgTyrTrpThrGlyProThrSerArgIleSerProGluAlaProValProValVal 41
Db 1 GATCGCTATTGGTATGCCGCCACTTGCCTATTTCACCGAAGCGCGTGCCTGGTT 60
QY 42 LysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAla 61
Db 61 AGGTAAATACCTGTGAGGAAGCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 62 SerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGly-IleAspAlaAla-ArgA 81
Db 121 TGTCTGTGGGCGAAGCGCGCTGTGGTGGCGCTGACGGGTATTGATGACGCGCGCGCGC- 179
QY 81 laLeuSerLysSerLeuAlaAspValAsnValLysCys-AspPheValSerValProThr 100
Db 180 -CTGAGCAAAACGCTGGCGGAGGTCAATGTGAAGTGGCGGACTTCGTTCTGTGCGGAGC 237
QY 101 HisProThrIleThrLysLeuArgValLeuSer-ArgAsnGlnGlnLeuIleArgLeuAs 120
Db 238 CATCCAGCATTAACCAAACTGGAGTACTATCTACGTATATCAGCAGCTCATTCGTTTGA- 296
QY 120 pPheGluGluGlyPheGluGlyValAspProGlnProLeuHisGlu 135
Db 297 -TTTGAAGAAGGCTTTGAGGAT---GACCGCAAGCGTTGCATGAG 338

RESULT 2

US-08-651-155B-186
; Sequence 186, Application US/08651155B
; Patent No. 6365401

GENERAL INFORMATION:

; APPLICANT: Mahan Dr., Michael J.
; APPLICANT: Conner Mr., Christopher P.
; APPLICANT: Hietchoff Mr., Douglas W.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; NUMBER OF SEQUENCES: 255
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chrismann, Bynum & Johnson, P.C.
; STREET: 1900 Fifteenth Street
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80302

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,155B
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:
; NAME: Petersen Mr., Steven C.
; REGISTRATION NUMBER: 36,238
; REFERENCE/DOCKET NUMBER: 17060.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/546-1300
; TELEFAX: 303/449-5426
; TELEX: ABAL475
; INFORMATION FOR SEQ ID NO: 186:

SEQUENCE CHARACTERISTICS:

; LENGTH: 294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-651-155B-186

Alignment Scores:

Pred. No.: 1,33e-20 Length: 294
Score: 263.00 Matches: 72
Percent Similarity: 79.59% Conservative: 6
Best Local Similarity: 73.47% Mismatches: 12
Query Match: 10.99% Indels: 10
DB: 4 Gaps: 5

US-09-912-020-325 (1-477) x US-08-651-155B-186 (1-294)

QY 267 IleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePhe 286
Db 2 ATCGCGCTGCTGGCGCGACCTG-GCCGCGGGAATACCTGGAAGAGCGGTGTTATTC 60
QY 287 AlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerProIle 306
Db 61 GCCTATCGCGCGCGCGCTAGTGTAGTAACCTCGGACGCTCAACGGTTTCCCTATT 120
QY 307 GluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGlu 326
Db 121 GAGCTGGAACACGACGAGTGGCGGAGC---GATACCGGCTTCGGCTTATGACCGAAGAG 176
QY 327 GluLeuLysLeuAlaAlaAlaAlaArgLysArgGlyGluLysValValMetThrAsn 346
Db 177 GAGTTGAGACAGCGCTGCCAGCGGTAAGTC-----GCGAGAAGTGTTCATGACCAAC 230
QY 347 GlyValPheAspIle---LeuHis-----AlaGlyHisValSerTyrLeu 360
Db 231 ---GCGTTCGATATCTGACGGCATATTATGACGCACTGGACCTATCGGACTTATA 281

RESULT 3

US-09-134-001C-1684
; Sequence 1684, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1684
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1684

Alignment Scores:

Pred. No.: 5.25e-09 Length: 948

Score: 165.50 Matches: 77
 Percent Similarity: 39.12% Conservative: 56
 Best Local Similarity: 22.65% Mismatches: 142
 Query Match: 6.92% Indels: 65
 DB: 4 Gaps: 13

US-09-912-020-325 (1-477) x US-09-134-001C-1684 (1-948)

QY 2 LysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMetLeu 21
 DB 7 AAAGTAGGTGAAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 66
 QY 22 AspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValVal 41
 DB 67 GATAAAATTTCTT-----AATGTTAAAGGTTTCCAAA---CCCGGTGAGACATTA 114
 QY 42 LysValAsnThrIleGluGluArgProGlyGly-----AlaAlaAsnValAlaMetAsn 59
 DB 115 CATATTAAACAGCTCAAAAGGAGTTTGGTGGGGCAAGGAGCCCAATCAAGCCATAGCA 174
 QY 60 IleAlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAla 79
 DB 175 GCTAGTAGATAGACACAGATCAACATTTATCAGTAAAGTTGGTAAAGATGCCAATGCT 234
 QY 80 ArgAlaLeuSerLysSerLeuAlaAspValAlaValLysCysAspPheValSerValPro 99
 DB 235 AACTTTATATGGAA-----GATTTCAAAAAGCAGGT 267
 QY 100 ThrHisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeu 119
 DB 268 ATTCAT-----ACAAATATATTTTAACTTCA 294
 QY 120 AspPheGluGlu-----GlyPheGluGlyValAspProGlnPro----- 132
 DB 295 GAAAGTGAAGAACTGGCGAAGCATTTATCACTGTTGATGAACGACGACAAATACGATT 354
 QY 133 ---LeuHisGluArgIleAsnGlnAlaLeuSer-----SerIleGly 145
 DB 355 CTGTGTTACGGTGGTGGCAATATGACATTAAGTGAACACTGATGTTGAGATGATGGAT 414
 QY 146 AlaLeuValLeuSerAspTyrAlaLysGlyAlaLeu-----AlaSerValGln 161
 DB 415 GCCTTTATGTCGACACTTTGTTAGCGCAGCTTGAAGTCCATTTGAGCGCATAGAA 474
 QY 162 GlnMetIleGlnLeuAlaGlyLysAlaGlyValProValLeuIleAspPro----- 178
 DB 475 CAAGCATTTAAATTCGCTGTAACAAATATCACTACTGATTAATACTCGCACCGCA 534
 QY 179 -----LysGlyThrAspPheGluArgTyrArgGlyAlaThr 190
 DB 535 ATTGAATTCGCTAAGTCACTTTTAGAGTTAACTGAT----- 570
 QY 191 LeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrGluGlu 210
 DB 571 ATAATTATTCACAAACGAAACGGAAGCAGAAATTTAATACAGGTATTTCAATCAATATGA 630
 QY 211 GluIleValGluArgGlyMetLysLeuIleAlaAspTyrGlnLeuSerAlaLeuVal 230
 DB 631 AGTGATATGAAGAACAACACACACATATTTTCGATTAGGTATATCTGCAGTATTAAT 690
 QY 231 ThrArgSerGlnGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetPro 250
 DB 691 ACTTTAGGGGAGCAAGCAGTATTGTCATATACAGACCAATACAAATG---ATTCCT 747
 QY 251 ThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeu 270
 DB 748 CGGTGTAATGAAGAACAATAGATACGACAGCAGCAGGAGATACATTTATAGGTGCTTT 807
 QY 271 AlaAlaThrLeuAlaAlaGly---AsnSerLeuGluGluAlaCysPhePheAlaAsnAla 289
 DB 808 TTAAGTAGGTAAATAAAGATTTGACAAATTTAGAAATCGGCTATTCGACTTCGAAATCAA 867
 QY 290 AlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerProIleGluLeuGlu 309

DB 868 GCGTCGCTCTTAACGCTACACGAAGAGGACACAGCTTCTATACCAACACGTAAAGAA 927
 RESULT 4
 US-08-920-812-19
 ; Sequence 19, Application US/08920812
 ; Patent No. 5763188
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uenara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/920,812
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,577
 ; FILING DATE: 27-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3756 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli
 ; STRAIN: Clinical Isolate EC-39
 ; US-08-920-812-19

Alignment Scores:
 Pred. No.: 1,86e-06 Length: 3796
 Score: 151.00 Matches: 179
 Percent Similarity: 39.62% Conservative: 47
 Best Local Similarity: 24.84% Mismatches: 156
 Query Match: 6.31% Indels: 37
 DB: 1 Gaps: 9

US-09-912-020-325 (1-477) x US-08-920-812-19 (1-3796)

QY 5 LeuProGluPheGluArgAlaGlyValMetValVal---GlyAspValMetLeuAspArg 23
 DB 2024 ATCCCGAATATGCAAAACGACGACGCGCTGTTGTTCTTGGCAGCATTAATGCTGACCAC 2083
 QY 24 TyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValLysVal 43
 DB 2084 ATT-----CTTAATCTTCAATCTTTTCTTACTCCAAGGAAACGTAACCGGTA 2131
 QY 44 AsnThrIleGluGluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSer 62

Qy 151 AspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArg-LysAl 170
Db 2455 -----CTGGAATACCACCTCGAAAGTGTGGTGGCAGCGGAGAAATCGCCCATCAAAAT 2508
Qy 170 aGlyValProValLeuIleAspProLysGlyThr-AspPhe---GluArgTyrArgGlyA 189
Db 2509 AAAAACTATCGTTCCTTAACCCCGTCCGGCTCGCGAATCTCTCAGCAAACTCTCGCGTG 2568
Qy 189 laThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrG 209
Db 2569 TGGACATATTACGCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2628
Qy 209 luGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuL 229
Db 2629 ATGATGAAGATCGACGAGCGGCGCGAGTACTTCTATCAAGAAAGGATATCCGTACTGTAC 2688
Qy 229 euValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHism 249
Db 2689 TGATTACTTTAGGAAGTCTGTGTGTATGGGTAGCGTGAATGGTGAAGGTACG---CGCG 2745
Qy 249 etProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyV 269
Db 2746 TTCCTGGATTCCGGGTGCAGGCTGTCGATACCATTCCTGCGGAGATACCTTTAAGCGTG 2805
Qy 269 alLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPheAlaAsnA 289
Db 2806 CGTTAATCAGCGCATGCTGGAAGAAACCAATTCAGAGGCGATTCGTTTGCCTCATG 2865
Qy 289 laAlaAlaGlyValValGlyLysLeuGlyLysLeuGlyThrSerThrValSerPro 305
Db 2866 CTGCGCGTGGATTCCGTAACACGTAAGCGGCAACACCTTCCTCGTACCG 2915

RESULT 6

US-08-921-177-19
; Sequence 19, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; CORRESPONDENCE ADDRESS: 25
; NUMBER OF SEQUENCES: 25
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3796 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-39
US-08-921-177-19
Alignment Scores:
Pred. No.: 1,86e-06 Length: 3796
Score: 151.00 Matches: 79
Percent similarity: 39.62% Conservative: 47
Best Local Similarity: 24.84% Mismatches: 156
Query Match: 6.31% Indels: 37
Gaps: 19
US-09-912-020-325 (1-477) x US-08-921-177-19 (1-3796)
Qy 5 LeuProGluPheGluArgAlaGlyValMetValVal---GlyAspValMetLeuAspArg 23
Db 2024 ATCCCGAATATGCAAAACGAGCGACCTCTGTTCTTGGCAGCATTAATGCTGACCAC 2083
Qy 24 TyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValValLysVal 43
Db 2084 ATT-----CTTAATCTTCAATCTTTCTCTACTCCAGCGGAAACGTAACCGGTA 2131
Qy 44 AsnThrIleGluGluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSer 62
Db 2132 ACCATATCAGGTTCGATTTTTCCTTACGGGTAAGCGGAAATCAGGCTGTGCTGCTGGCGCT 2190
Qy 63 LeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeu 82
Db 2191 AGCGGTGCGAATATCGCGTTTATTTCCTTACGGGTGATGACGACGATTGTCGAGACGTT 2250
Qy 83 SerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThrHisPro 102
Db 2251 CGCCAGCAGCTCGCCACTGATAACATT----- 2277
Qy 103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGlu 122
Db 2278 GATATTACTCCGTCGCGTGATCAAGCGCAATCAACAGGTGTGCGCTGTATTGTT 2337
Qy 123 GluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSer 142
Db 2338 AATGCG---GAAGTGAGAAATCTCATGCTATTCATCCGCGGCTAATGCTGCCCTTCC 2394
Qy 143 -----SerIleGlyAlaLeuValLeuSer 150
Db 2395 CGCGCGCTGTGGAAAGCGCAACGTGAGCGTATTGCCAAGCGCTCAGCATTAATATGCAG 2454
Qy 151 AspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArg-LysAl 170
Db 2455 -----CTGGAATACCACCTCGAAAGTGTGGTGGCAGCGGAGAAATCGCCCATCAAAAT 2508
Qy 170 aGlyValProValLeuIleAspProLysGlyThr-AspPhe---GluArgTyrArgGlyA 189
Db 2509 AAAAACTATCGTTCCTTAACCCCGTCCGGCTCGCGAATCTCTCAGCAAACTCTCGCGTG 2568
Qy 189 laThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrG 209
Db 2569 TGGACATATTACGCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2628
Qy 209 luGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuL 229
Db 2629 ATGATGAAGATCGACGAGCGGCGCGAGTACTTCTATCAAGAAAGGATATCCGTACTGTAC 2688
Qy 229 euValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHism 249
Db 2689 TGATTACTTTAGGAAGTCTGTGTGTATGGGTAGCGTGAATGGTGAAGGTACG---CGCG 2745
Qy 249 etProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyV 269
Db 2746 TTCCTGGATTCCGGGTGCAGGCTGTCGATACCATTCCTGCGGAGATACCTTTAAGCGTG 2805
Qy 269 alLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPheAlaAsnA 289
Db 2806 CGTTAATCAGCGCATGCTGGAAGAAACCAATTCAGAGGCGATTCGTTTGCCTCATG 2865
Qy 289 laAlaAlaGlyValValGlyLysLeuGlyLysLeuGlyThrSerThrValSerPro 305
Db 2866 CTGCGCGTGGATTCCGTAACACGTAAGCGGCAACACCTTCCTCGTACCG 2915

Db 2746 TTCCTGATCCGGGTGAGCGTGTGCGATACCATTCCTCCGGGAGATACCTTTAAACGGTG 2805
QY 269 allLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhePheAlaAsnA 289
Db 2806 CGTTATATCAGGATTCGTGGAGAAACCATTCGACAGGCGATTCGTTTGGCCCATG 2865
QY 289 laLaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 2866 CTGCCGCTGCGATTGCGGTAAACAGTAAAGCGCAACACTTCCTCGTACCG 2915
RESULT 7
US-08-362-577C-19
; Sequence 19, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,577C
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical isolate EC-39
; US-08-362-577C-19
Alignment Scores:
Pred. No.: 1,86e-06 Length: 3796
Score: 151.00 Matches: 79
Percent Similarity: 39.62% Conservative: 47
Best Local Similarity: 24.84% Mismatches: 156
Query Match: 6.31% Indels: 37
DB: 1 Gaps: 9
US-09-912-020-325 (1-477) x US-08-362-577C-19 (1-3796)

Db 2084 ATT-----CTTAATCTTCAATCTTTTCCTACTTCCAGCGCAACAGTAAACGGTA 2131
QY 44 AsnThrIleGluGluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSer 62
Db 2132 ACCACTATCAGTTGATTTT-GGCGCAAGCGCGAATCAGGCTGTGGCTGTGGCGGT 2190
QY 63 LeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeu 82
Db 2191 AGCGGTGCGAATATCGCTTTATTGCTGTGACGGGTGATGACAGCATTTGTTGAGAGCGTT 2250
QY 83 SerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProHisPro 102
Db 2251 CGCCAGCAGCTCGCCACTGATAACAT----- 2277
QY 103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGlu 122
Db 2278 GATATTACTCCGGTCAGCGTGATCAAGGCGAATCAACAGGTGTGGCGGTGATTTTGT 2337
QY 123 GluGlyPheGluGlyValAspProGlnProLeuHisGlnArgIleAsnGlnAlaLeuSer 142
Db 2338 AATGGC---GAAGGTGAGAATGTATCGGTATTCATCGCGCGCGTAATGCTGCCCTTCC 2394
QY 143 -----SerIleGlyAlaLeuValLeuSer 150
Db 2395 CGGCGCTGTGGAGCGCAACGTGAGCGTATTGCGAAGCGGTACGATTTATTATGTCAG 2454
QY 151 AspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArg-LysAl 170
Db 2455 -----CTGAATCACCACCTCGAAGTGTGATGCGACGCGCGAAGTTCCTCGAACTCTCGCGTG 2508
QY 170 aglyValProValLeuIleAspProLysGlyThr-AspPhe---GluArgTyrArgGlyA 189
Db 2509 AAAAATATCGTTTAAACCGCTCGCGCTCGCGAAGTTCCTCGAACTCTCGCGTG 2568
QY 189 laThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrG 209
Db 2569 TGGACATTATTACGCCAAACGAAACGAGCAGCAAGCTCACCGGTATTTCGTGTTGAAA 2628
QY 209 luGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuL 229
Db 2629 ATGATGAAGATGACGAGCGCGCGCGAGTACTTCTGATAAAAGGTATCCGTACTGTAC 2688
QY 229 euValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisM 249
Db 2689 TGATTACTTTAGGAAGTCTGCTGTGATGGCTAGCGTGAATGCTGAAGGTGAG---CGCG 2745
QY 249 etProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyV 269
Db 2746 TTCCTGATTCCGGGTGACGGCTGTGATACCATTCGTCGGGAGATACCTTTAAACGGTG 2805
QY 269 allLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhePheAlaAsnA 289
Db 2806 CGTTAATCAGGCGATTCCTCGAAGAAACCATTCGACAGGCGGATTCGTTTGGCCCATG 2865
QY 289 laLaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 2866 CTGCCGCTGCGATTGCGGTAAACAGTAAAGCGCAACACTTCCTCGTACCG 2915
RESULT 8
US-08-920-828-19
; Sequence 19, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago

```
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 3796 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: Clinical Isolate EC-39
US-08-920-828-19

Alignment Scores:
Pred. No.: 1,86e-06 Length: 3796
Score: 151.00 Matches: 79
Percent Similarity: 39.62% Conservative: 47
Best Local Similarity: 24.84% Mismatches: 156
Query Match: 6.31% Indels: 37
DB: 2 Gaps: 9

US-09-912-020-325 (1-477) x US-08-920-828-19 (1-3796)

Qy 5 LeuProGluPheGluArgAlaGlyValMetValVal---GlyAspValMetLeuAspArg 23
Db 2024 ATCCGGAATATGCAAAAGCGGAGCGCTCTGTTGTCGACGCAATTAATGCTGACAC 2083

Qy 24 TyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValValLysVal 43
Db 2084 ATT-----CITTAATCTCAATCTTTCTTCTACTCCAGCGAAACGTAACCGGTA 2131

Qy 44 AsnThrIleGluGluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSer 62
Db 2132 ACCATATCAGGTGTCATTT-GGCGGAAAGCGGCAATCAGGCTGTGCTGCTGGCGGT 2190

Qy 63 LeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeu 82
Db 2191 AGCGGTGCGAATATCGCGTTTATTGCCGTACGGGTGATGACAGCATTTGGTGAGAGCGTT 2250

Qy 83 SerIleSerLeuAlaAspValAsnValLysCysAspPheValSerValProThrHisPro 102
Db 2251 CGCCAGCAGCTCGCCATCATTAACATT----- 2277

Qy 103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGlu 122
Db 2278 GATATTACTCGCGTCAGCGTGATCAAAAGCGCAATCAACAGGTGTGGCTGATTTTGT 2337

Qy 123 GluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSer 142
Db 2338 AATGCG---GAAGGTGAAGATGTATCATCGGTATTATCGCGCGCTAATGCTGCCCTTCC 2394

Qy 143 -----SerIleGlyAlaLeuValLeuSer 150
Db 2395 CCGCGCGTGGTGAAGCGCAACGTCAGCGTATTGGCAACGCGTCAGCATTTATTATGCAG 2454

Qy 151 AspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArg-LysAl 170
Db 2455 -----CTGGAATCACCACCTCGAAAGTGTGATGGCAGCGCGGCAAAATCCCATCAAAAT 2508

Qy 170 aGlyValProValLeuIleAspProLysGlyThr-AspPhe---GluArgTyrArgGlyA 189
Db 2509 AAAAATATATCGTTCGTTAACCCGCTCCGGCTCGCGAATCTCTCAGCAACTCTCGCGTG 2568

Qy 189 laThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrG 209
Db 2569 TGGACATTATTACGCCAAACGGAAGCAGAGAAAGCTCACCGGTATTCTGTTGAAA 2628

Qy 209 luGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuL 229
Db 2629 ATGATGAAGATGCAGCGAAGCGCGCAGGTACTTTCATGAAAAAGGTATCCGTACTGTAC 2688

Qy 229 euValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisM 249
Db 2689 TGATTACTTTAGGAAGTCGTGTGTGGCTAGCGTAGCGTGAATGGTGAAGGTACAG---CGCG 2745

Qy 249 etProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyV 269
Db 2746 TTCCTGGATTCCGGGTGCAGGCTGTCGATACCATTCGTCGCGGAGATACCTTTACGGTG 2805

Qy 269 alLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnA 289
Db 2806 CGTTAATCAGCGCATTCGTGGAAGAAAAACCATTCGACAGAGCGGATTCGTTTGCCTCATG 2865

Qy 289 laAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 2866 CTGCGCGTGGCATTCGCGTAACACGTAAGGCGCACAAACCTTCCCTACCG 2915

RESULT 9
US-09-134-001C-2723
; Sequence 2723, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2723
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2723

Alignment Scores:
Pred. No.: 7.5e-08 Length: 414
Score: 150.00 Matches: 47
Percent Similarity: 54.68% Conservative: 29
Best Local Similarity: 33.81% Mismatches: 49
Query Match: 6.27% Indels: 14
DB: 4 Gaps: 6

US-09-912-020-325 (1-477) x US-09-134-001C-2723 (1-414)

Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 19 AAAAGAGTCATTACTTATGGTACATGACTTGTCTTATGTTATGTCATTTGAATTACTT 78

Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
```

Db	79	AGAAGAGCTCGTGAGATGGGGATTATCTATCGTGGCGCTTCTACTGAT---	GAATTC	135
	::	
QY	381	LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu	400	
	::	
Db	136	AACCAATCAAA--AACRAAAATCATATTATGATTATGACACCGTAAGATGATGTTA	192	
	::	
QY	401	GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu	420	
	::	
Db	193	GAGCTATTAGATACGTAGACTTAGTTATCCCT--GAAGAAGGATGGGACAGAAAGAA	249	
	::	
QY	421	-----IleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTyrIysProGlu	438	
	::	
Db	250	AAAGATGTGTGATCGTTTGTGATGATGATTGTGTGATGGTCATGATGGAA-----	303	
	::	
QY	439	GluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPhe	458	
Db	304	-----GGGAGATTGACTCCCTTAAGATAGTGTGAAGTCATTATCTTAACCCG	354	
QY	459	GluAspGlyCysSerThrThrAsnIleIleLysLysIleGlnAspLysLysGly	477	
	::	
Db	355	ACAGAAGGTATCTCAACTACC-----AAATCAAGCAAGAAATTATACGGA	399	
	::	

RESULT 10

```

US-09-453-702B-149/c
; Sequence 149, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Nicole T.
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinkney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 149:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12804
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 149:
US-09-453-702B-149

```

Score:	146.50	Matches:	81
Percent Similarity:	39.04%	Conservative:	49
Best Local Similarity:	24.33%	Mismatches:	136
Query Match:	6.12%	Indels:	67
DB:	4	Gaps:	15
US-09-912-020-325 (1-477) x US-09-453-702B-149 (1-12804)			
Qy	11	AlaGlyValMetValValGlyAspValMetLeuAspArgTyrTyrGlyProThrSer	30
Db	9502	GCCAAAGTATGGGTTTAGGGGATCGGTCGPAGAT-----	9867
Qy	31	ArgIleSerProGluAlaProValProValValLysValAsnThrIleGluGluArgPro	50
Db	9866	--CTCTTGCCAGAAATACAGCGGGGGTACTGCGCTGT-----	9828
Qy	51	GlyGlyAla---AlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeu	69
Db	9827	GGCGGGCGCCAGCTAACGTTGCGGTGGGAATCGCCAGATTAGCGGGAACAAAGTGGGTTT	9768
Qy	70	ValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeuAlaAspVal	89
Db	9767	ATAGGTGGGTGGGGGATGATCCTTTGGTGTCATTATGCAAAAGACGCTGCTAACTGAG	9708
Qy	90	AsnValLysCysAspPheValSerValProThr---HisProThrIleThrLysLeuArg	108
Db	9707	GGAGTCGATATCATGTTATCTGAAGCAAGATGAATGGCACCGGACATTCACGGTGCTT---	9651
Qy	109	ValLeuSerArgAsnGlnGln-----LeuIleArgLeuAspPhe	121
Db	9650	--GTGATCTGAACGATCAAGGGGACGTTCAATTTAGTTTGGTCCGCCGCCAGTGCC	9594
Qy	122	GluGluGlyPheGlyValAsp-----ProGlnProLeuHisGlu	135
Db	9593	GATCTTTTTTAGAGCAGACACTGGCCCTCTGGCGACATGGCGAATGGTTACAT---	9537
Qy	136	ArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGly	155
Db	9536	-----CTCTGTTCAAATT---GGTTGTCTGCGGAGCTTTCGCGTACCAGC	9495
Qy	156	AlaLeuAlaSerValGlnGlnMetIleGlnIleAlaArgLysAlaGlyValProValLeu	175
Db	9494	GCATTTACTCGGATGACGGAGATC-----CGGCATGCGGAGGTTTGTCAGC	9447
Qy	176	IleAspProLysGlyThrAsp-----	182
Db	9446	TTCGATCCCAATAATTGCTGAAGATCTATGGCAAGCAGGACATTGCTCCGCTGTGTTTG	9387
Qy	183	PheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaVal	202
Db	9386	CGGCGGCGGTACAACTGGCGGATGTCGTCAAGCTCTCGGAAGAGAATGGCGACTTATC	9327
Qy	203	ValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAsp	222
Db	9326	AGTGGAAAAACACAGAACCATCGGGATATA-----TGGCGCCCTGGCAAAAGAG	9279
Qy	223	TyrGluLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnPro	242
Db	9278	TATGAGATCCCAATGCTGTGTGTGACTAAAGTGTCAGAAGGGGTGGTGTCTGTATTATCGA	9219
Qy	243	GlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGlyAla	262
Db	9218	GGACAAGTTTCAC---CATTTTGTCTGGAATGTCTGTGGATTGTGTGATAGCAGGGGCG	9162
Qy	263	GlyAspThr---ValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu	281
Db	9161	GGAGATCGGTTCGTTGCCGGGTACTCACAGCTGTCTCTACGGGATATTACTACAGAT	9102
Qy	282	Glu-----AlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGly	296
Db	9101	GAGAGAGAAATGCGACAAATATTATCATCTGCTCAACGTTGCGGAGCGCTTGCAGTAAAG	9042
Qy	297	LysLeuGlyThrSerThrValSerProIleGluLeuGlu	309

Db 9041 GCGAAGGGGCAATGACAGCGCTGCCATGTCCACAAGAA 9003

RESULT 11

```

US-09-199-637A-280
; Sequence 280, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 280
; LENGTH: 2580
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-280

```

Alignment Scores:		
Pred. No.:	2,45e-05	Length:
Score:	138.50	Matches:
Percent Similarity:	39.38%	Conservative:
Best Local Similarity:	25.15%	Mismatches:
Query Match:	5.73%	Indels:
DB:	4	Gaps:
		28
		3580

Qy	39	ProValValLysValAsnThrIleGluGluArgProGlyGlyAla	-----	53
Db	106	CCGCTGCACCTGTTTCCGCCCTGCTCGCAGCAGCAAGCGGTTCGATCAAGCCCTCGTG	165	
Qy	54	AlaAsnValAlaMetAsnIleAlaSerLeu	-----GlyAlaAsnAlaArgLeuValGly	71
Db	166	ATGCAGTTCGGTTCGATATCGCGCCCTGCGCAGCGGCTCAACAAAGACTCGACGGC	225	
Qy	72	LeuThrGlyIle	-----AspAlaAlaAlaArgAla	81
Db	226	CTGCCGAAGATCCAGAGCCCGACCGCGACGTGAACCTGTCCAGGATCTCGCAGCGCTG	285	
Qy	82	LeuSerLysSer	-----LeuAlaAspValAsnValLysCysaspPhe	95
Db	286	CTCAACCGAGCTGACCGCTGGCCCGCCAGCAGAGGGCGACAGTTCACTCCAGCAGGNG	345	
Qy	96	ValSerValProThrHisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGln	115	
Db	346	GTATTGTGCGCGCGATGGACGAGAACACCAAGSCTCGGCAAGCTGCTGTGCGCCAGGCG	405	
Qy	116	LeuIleArgLeuaspPheGlu	-----GluGlyPheGluGlyValasp	129
		:::	:::	
Db	406	GTGTGCGCGAAGCGCTGGAGATGCGTGCCAACTCGTGGCGCGCAAGCGGTGAAC	465	
Qy	130	ProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeu	149	
Db	466	--GACCCGAACCTCGAGGAGTCGCCCGCAGCGCTGGCAAGTACACCTCCACATGACC	522	
Qy	150	SerAspTyrAlaLysGlyAlaLeu	-----AlaSerValGlnGln	162
		:::	:::	
Db	523	AAGCGCGCCGGAAGGCAAGCTCGACCGGTTGATCGTGGCGACGACGAGATCCGCCGG	582	

Qy	163	MetIleGlnLeu-----AlaArgLysAlaGlyValProValLeuIleAspProLysGly	180
		:::	:::
Db	583	ACCATCCAGGTCTCGACGGCGGCCAACAAACACCCGGTGCCTGATCGCGGAACCCGGC	642
Qy	181	ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSer-----	197
		:::	:::
Db	643	GTC-----GGCAAGACCGGCATCTGCGAGCGCGCTGCGCCACCGCATC	684
Qy	198	-----GluPheGlu	200
		:::	:::
Db	685	ATCAACGGCCAAAGTCGGACGGCCCTCAAGGACAAGCGCTGCTGGCCCTGGACATGGG	744
Qy	201	AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle	220
		:::	:::
Db	745	CGCGTATCCCGGTCCGAAGTTCCGCGCGAGTTCGAGGAACGC-----	789
Qy	221	AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu	240
		:::	:::
Db	790	-----CTGAAGCGGTCTCT-----AACGAACCTGGGCAG-----	819
Qy	241	GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGlnValTyrAspValThr	260
		:::	:::
Db	820	CAGGAAGCGGGTCATCTCTGTTTCATC-----GACGAACCTGCACACCATGGTC	867
Qy	261	GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu	280
		:::	:::
Db	868	GGCCCGGCACAGCG-----GAAGTGCATCGGACGCGCGCAACATGCTC	912
Qy	281	GluGluAlaCysPhe-PheAlaAsnAlaAlaGlyValValGlyLysLeuGlyTh	300
		::: ::::: ::::: ::::: ::::: :::::	
Db	913	AAGCGCGCTCGCGCGCGCGAGCTGCACCTCGCTGGTGTACTA-----CC	960
Qy	300	rSerThr-ValSerProIleGluLeuGluAsnAlaValArgLysArg--AlaAspThrG	319
		:::	:::
Db	961	CTCAGCAGCATCGCCAGTACATCGAGAAGGATCGCGCGCTGGAGCGCGCTCCACAAG	1020
Qy	319	lyPheGlyValMetThrGlu-GluGluLeuLysLeuAlaValAlaAlaAlaArgLysArg	338
		:::	:::
Db	1021	GTGCTGTGCACGAACCGACGAGGAAGACACCATCGCATCTCCGTGGCTCAAGAA	1080
Qy	339	GlyGluLys-----ValValMetThrAsnGlyValPheAspIleLeuHisAla	354
		:::	:::
Db	1081	CGCTATGAATGCAACACCGGGTGACATCAACCGACGGCGG-----ATCATCGCCGCG	1134
Qy	355	GlyHisValSer-----TyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuIleVal	372
		:::	:::
Db	1135	GCCAAAGCTGTGCAACCGCTACATCCAGT---CGGAACCTGCCGACCAAGCCATCGAC	1191
Qy	373	AlaValAsnSerAspAlaSerThrLysArgLeuLysGlyAspSerArgPro-----Val	390
		:::	:::
Db	1192	CTGATCACCAGGCGCGCCATCCGCATCGATGGAGATCGACTCCAGCCGCGAGAACTG	1251
Qy	391	AsnProLeuGluGlnArgMetIleValLeu-----GlyAlaLeuGluAlaVal	406
		:::	:::
Db	1252	GATCGTCTCCACCGTCGCTGTATCCAGTCAAGATCGAGCCGCGAGGCGCTGAAGAAGAA	1311
Qy	407	AspTrpValValSerPheGluGluAspThrProGlnArgLeuIleAlaGlyIleLeuPro	426
		:::	:::
Db	1312	GAC-----GACGAAGCCACGAGAGCGGCTGCCAAG-----CTGGAG	1350
Qy	427	AspLeuLeuValLysGlyGlyAspTyrLysProGluIleAlaGlySerLysGluVal	446
		:::	:::
Db	1351	GAGGATATCTGTAAG-----CTCAGAGCGGAATACGCGGACCTCGAGGAGATC	1398
Qy	447	Trp---AlaAsnGlyGlyGluValLeuValLeuAsnPheGluAspGlyCysSerThrThr	465
		:::	:::
Db	1399	TGGAAGTCCGAGAGGCGAGGTGCAG-----GGCTCGCGC	1434
Qy	466	AsnIleIleLysLysIleGlnAlaAspLysLys	476
		:::	:::
Db	1435	CAGATCCAGCAGGAATCGAGCAGCAAGCAG	1467
		:::	:::
			RESULT 12

US-09-199-637A-272/c
; Sequence 272, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-272

Alignment Scores:
Pred. No.: 3,07e-05 Length: 2970
Score: 138.50 Matches: 129
Percent Similarity: 39.38% Conservative: 73
Best Local Similarity: 25.15% Mismatches: 179
Query Match: 5.79% Indels: 132
DB: 4 Gaps: 28

US-09-912-020-325 (1-477) x US-09-199-637A-272 (1-2970)

QY 39 ProValValLysValAsnThrIleGluArgProGlyGlyAla----- 53
| | | | | : : : : : | | | | : : : : :
DB 2859 CCGGTGCACCTGCTTCCCGCCCTGCTCGAGCAGCAGCGGTTCGATCAAGCCCTGCTG 2800
QY 54 AlaAsnValAlaMetAsnIleAlaSerLeu-----GlyAlaAsnAlaArgLeuValGly 71
| | | | | : : : : : | | | | : : : : :
DB 2799 ATCAGGTGGCTTCGATATCGCGCCCTCGCAGCGGCTCAACAAGAACTCGACGG 2740
QY 72 LeuThrGlyIle-----AspAlaAlaArgAla 81
| | | | | : : : : : | | | | : : : : :
DB 2739 CTCGCCAAGATCCAGAGCCGACCGCGGACGTGAACCTGTCGAGGATCTCGCACCCCTG 2680
QY 82 LeuSerLysSer-----LeuAlaAspValAsnValLysCysAspPhe 95
| | | | | : : : : : | | | | : : : : :
DB 2679 CTCACACAGGCTGACCGCTGGCCAGCAGAGGGCGACCATTCATCTCCAGCGAGCTG 2620
QY 96 ValSerValProThrHisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGln 115
| | | | | : : : : : | | | | : : : : :
DB 2619 GTATTGTCGCCGCGATGACGAGAACACACAGGCTCGGCAAGCTGCTCGCGCCAGGGC 2560
QY 116 LeuIleArgLeuAspPheGlu-----GluGlyPheGluGlyValAsp 129
| | | | | : : : : : | | | | : : : : :
DB 2559 GTCTCGCGCAAGCGCTGGAGAAATGCGGTGGCCAACTCGGTGGCGGGAAGCGGTGAAC 2500
QY 130 ProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeu 149
| | | | | : : : : : | | | | : : : : :
DB 2499 ---GACCCGAACCTCGAGGATGCGGCCAGCGCTGACAAAGTACACCGTCGACATGACC 2443
QY 150 SerAspTyrAlaLysGlyAlaLeu-----AlaSerValGlnGln 162
| | | | | : : : : : | | | | : : : : :
DB 2442 AAGCGCGCGGAGGAGGCAAGCTCGACCGGCTGACCGGTCGCGACGACGAGATCCGCGG 2383
QY 163 MetIleGlnLeu-----AlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
| | | | | : : : : : | | | | : : : : :
DB 2382 ACCATCCAGGTCTCGACGGCGGACCAAGAACACCCCGTGTGATCGCGCAACCCGGC 2323

QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSer----- 197
| | | | | : : : : : | | | | : : : : :
DB 2322 GTC-----GGCAAGACCGGCATCTCGAGCGCTGCGCCAGCGCATC 2281
QY 198 -----GluPheGlu 200
| | | | | : : : : : | | | | : : : : :
DB 2280 ATCAAGGCGAAGTCCCGACGCGCTCAAGGACAAGCGCTGCTGGCCCTGGACATGGGG 2221
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
| | | | | : : : : : | | | | : : : : :
DB 2220 GCGCTGATCCCGGTCGCAAGTCCCGCGGAGTTCCGAGGAAGCG----- 2176
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
| | | | | : : : : : | | | | : : : : :
DB 2175 -----CTGAAGCGGTCTC-----AAGCAACTGGCAAG----- 2146
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValValValThr 260
| | | | | : : : : : | | | | : : : : :
DB 2145 CAGGAAGCGCGGTCTCATCTGTTCATC-----GACGAATGACACCATGGTC 2098
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaThrLeuAlaAlaGlyAsnSerLeu 280
| | | | | : : : : : | | | | : : : : :
DB 2097 GCGCGCGCAAGCG-----GAAGTGCCATGGACCGCGCAACATGCTC 2053
QY 281 GluGluAlaCysPhe-PheAlaAsnAlaAlaAlaGlyValValValIleLysLeuGlyTh 300
| | | | | : : : : : | | | | : : : : :
DB 2052 AAGCGGCTCTGGCGCGCGGAGTGCACCTGGTGGTGTCTACTA-----CC 2005
QY 300 rSerThr-ValSerProIleGluLeuGluAsnAlaValArgGlyArg---AlaAspThrG 319
| | | | | : : : : : | | | | : : : : :
DB 2004 CTCGACGAGTATCGCCAGTACATCGAGAAGGATCGCGCGTGGAGCGCGCTTCCAGAA 1945
QY 319 lypheGlyValMetThrGlu-GluGluLeuLysLeuAlaValAlaAlaAlaArgLysArg 338
| | | | | : : : : : | | | | : : : : :
DB 1944 GTGCTGTGACGACCGGAGGAGACACCATCGCATCTCTCGTGGCTCAAGGAA 1885
QY 339 GlyGluLys-----ValValMetThrAsnGlyValPheAspIleLeuHisAla 354
| | | | | : : : : : | | | | : : : : :
DB 1884 CGCTATGAGTGACACCGCGGTGAGCATCACGACGCGCG-----ATCATCGCGCG 1831
QY 355 GlyHisValSer-----TyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuIleVal 372
| | | | | : : : : : | | | | : : : : :
DB 1830 GCCAAGCTGCGACCGCTACATCACCGAT---CGCAACTGCGCCGACAGGCGCATCGAC 1774
QY 373 AlaValAsnSerAspAlaSerThrLysArgLeuLysGlyAspSerArgPro-----Val 390
| | | | | : : : : : | | | | : : : : :
DB 1773 CTGATCGAGCGCGCCCGACCGCATCCGATGGAGATCGACTCAGCGGAGGAACTG 1714
QY 391 AsnProLeuGluGlnArgMetIleValLeu-----GlyAlaLeuGluAlaVal 406
| | | | | : : : : : | | | | : : : : :
DB 1713 GATCGTCTCGACGCTCGCTGATCCAGCTGAAGATCGAGCGCGCAACGCTGAAGAGGAA 1654
QY 407 AspTrpValValSerPheGluGluAspThrProGlnArgLeuIleAlaGlyIleLeuPro 426
| | | | | : : : : : | | | | : : : : :
DB 1653 GAC-----GACGAAGCCACCGAAGCGCTGCGCAAG-----CTGGAG 1615
QY 427 AspLeuLeuValLysGlyGlyAspTyrLysProGluGluIleAlaGlySerLysLys 446
| | | | | : : : : : | | | | : : : : :
DB 1614 GAGGATATCGTCAAG-----CTCGAGCGGAGTACCGCCACCTCGAGGAGATC 1567
QY 447 Trp---AlaAsnGlyGlyGluValValLeuAsnPheGluAspGlyCysSerThrThr 465
| | | | | : : : : : | | | | : : : : :
DB 1566 TGAAGTCCGAGAAGCGCGAGGTGCAG-----GGCTCGGCG 1531
QY 466 AsnIleIleLysLysIleGlnGlnAspLysLys 476
| | | | | : : : : : | | | | : : : : :
DB 1530 CAGATCCAGCAAGATCGAGCGCAAGCAG 1498

RESULT 13

US-09-221-017B-757
; Sequence 757, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 NUMBER OF SEQUENCES: 1120
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018

COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 FILING DATE: 23-DEC-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1182
 FILING DATE: 31-DEC-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1546
 FILING DATE: 30-JAN-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP2911
 FILING DATE: 09-APR-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU98/01023
 FILING DATE: 10-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Monroy, Gladys H
 REGISTRATION NUMBER: 32,430
 REFERENCE/DOCKET NUMBER: 27340-20021.00
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141

INFORMATION FOR SEQ ID NO: 757:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3058 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:
 ORGANISM: PORPHYROMONAS GINGIVALIS
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1...3058
 US-09-221-017B-757

Alignment Scores:
 Pred. No.: 3,21e-05 Length: 3058
 Score: 138.50 Matches: 95

Percent Similarity: 35.01% Conservative: 72
 Best Local Similarity: 19.92% Mismatches: 211
 Query Match: 5.79% Indels: 99
 DB: 4 Gaps: 15

US-09-912-020-325 (1-477) x US-09-221-017B-757 (1-3058)

QY 43 ValAsnThrIleGluGluArgProGlyAlaAlaAsnValAlaMetAsnIleAlaSer 62
 DB 519 GTAAATGTCATCCCGAA-----GAGCCAAATGGAGATGGCGTTTCTCT 566

QY 63 LeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeu 82
 DB 567 CCTAGGCGGAGCC-----TATGGATAAAAGCTCTGGACATGGCCATCAAGCGTG 620

QY 83 SerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThrHisPro 102
 DB 621 AACCAAGCGATCGGATGTG-----CTGGTGACCATCGGCATCAACAA 665
 QY 103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGlu 122
 DB 666 GCAGCCATCGCGCAGGACATCTTTCCATACAAAGGCCATACGCAATACCTGCAGGCCAA 725
 QY 123 GluGlyPheGluGlyValAspPro----- 130
 DB 726 GCCGCATAGAGGCGGAAGATCGCTGATGCTCGCGCAGGATGAAGCTCCGACAGCT 785
 QY 131 ---GlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeu 149
 DB 786 TTGGTAACGATGCAGCGCGATAGCAGCGTCCCTATCGCTTATCTACTCGGAGCAATC 845
 QY 150 SerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLys 169
 DB 846 CTGGACAAGCTGAAGGCTTTTCGATCGCAGCTCCGCGATGGATTTCTGGTATAGTACGTCG 905
 QY 170 AlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArgTyrArgGlyAla 189
 DB 906 CGGATAGCGCTCATGGCTCTGAATCCGATCGCGCGC----- 944
 QY 190 ThrLeuLeuThrProAsnLeuSerGluPheGluAlaValAlaValGlyLysCysLysThrGlu 209
 DB 945 -----ACGGACTGATCGC-----ACCGAG 965
 QY 210 Glu-----GluIleValGluArgGlyMetLysLeuIle 220
 DB 966 GAAAGCGACAGCATCGCGCGGTGCAAGAGGAGCAAGACAGGATTTGTAGTGTTC 1025
 QY 221 AlaAspTyr-----GluLeuSerAlaLeu 228
 DB 1026 GGCCTTATGCGCTGACGGTTTTTGGGGTGGGATGATGGCTGTCATTTTCGATGGCAT 1085
 QY 229 LeuValThrArgSerGluGlnGlyMetSerLeuLeuGln----- 241
 DB 1086 CTGTCCATGATCATCATCAGGGGCTGATCCCTTTCAAGACTCTCTGCATGGACAGGG 1145
 QY 242 -----ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyr 257
 DB 1146 GTCAATGTGACGCGCAGGGCTTTCCATCGTACGCACCTTCGCGGATCATGGTACAGGCTTC 1205
 QY 258 AspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGly 277
 DB 1206 GACATCGTCGCAAGGCGGAG-----GCTTCGCCC 1235
 QY 278 AsnSerLeuGluGluAlaCysPheAlaAsnAlaAlaGlyValValValGlyLys 297
 DB 1236 GATTCCTTCGCGCGCAGCCATCTAT-----CAGCGCATCGACATTTATCGTTCGCGG 1286
 QY 298 LeuGly-----ThrSerThrValSerProIleGluLeuGluAsnAlaValArgGlyArg 315
 DB 1287 GCATCGTGGCGATCGGCTACGCGCAATCTCTCCGAAAGAGTTATTTCGAGCGGCGCAAT 1346
 QY 316 AlaAspThrGlyPheGlyValMetThrGluGluGluLeuLysLeuAlaValAlaAla 335
 DB 1347 GACATGAGAAATGCGCGCAACAGAGATGAGCATTAATTTAATACCCCTATAGACTCT 1406
 QY 336 ArgLysArgGlyGluLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGly 355
 DB 1407 ATGTTTAAGSATAGACTATCGTTTACACATCGGCGACATTCGATGTTCACATACAC 1466
 QY 356 HisValSerTyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsn 375
 DB 1467 CATTTCCGCGATGATCAACTATGCGCGCAGTTTGGCAGACATCTTATCGTTGGCGTCAGC 1526
 QY 376 SerAspAlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluGln 395
 DB 1527 ACGGAGCAACTGCTGCTTCATACAAG-----GCGCCGCCCATTTATCTCTTTATCGAG 1580
 QY 396 ArgMetIleValLeuGlyAlaLeuGluAlaValAspThrValValSerPheGluGluAsp 415

Db 24880 GAC-----GACGAAGCCACCAGGAAGCGCTGGCCAAG-----CTGGAG 24842
Qy 427 AspLeuLeuValLysGlyGlyAspTyrLysProGluGluIleAlaGlySerLysGluVal 446
Db 24841 GAGGATTCGTCAG-----CTCGAGGCGGAATACGCGGACCTCGAGGAGATC 24794
Qy 447 Trp---AlaAsnGlyGlyGluValLeuValLeuAsnPheGluAspGlyCysSerThrThr 465
Db 24793 TGGAAAGTCGAGAGCGCGAGGTGCAG-----GGCTCGGCG 24758
Qy 466 AsnIleLeuLysLysIleGlnAspLysLys 476
Db 24757 CAGATCCAGCAGAGATCGAGCGCAAGCAG 24725

RESULT 15

US-09-134-001C-2746
; Sequence 2746, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2746
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2746

Alignment Scores:

Pred. No.: 6.55e-06 Length: 939
Score: 137.50 Matches: 58
Percent Similarity: 43.28% Conservative: 58
Best Local Similarity: 21.64% Mismatches: 411
Query Match: 5.75% Indels: 41
DB: 4 Gaps: 12

US-09-912-020-325 (1-477) x US-09-134-001C-2746 (1-939)

Qy 51 GlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeuVal 70
Db 127 GGTAAAGGAATCAATGTATCAAGAGTACTTAAAGCTTTAGATGTTGATCTACAGCCCTTA 186
Qy 71 GlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeuAlaAspValAsn 90
Db 187 GGATTTTCAGGA---GGTTTCCCTGGAGATTTATTGTCTCAACACTTGAGCAGCAGTAAT 243
Qy 91 ValLysCysAspPheValSerValProThrHisProThrIleThrLysLeuArgValLeu 110
Db 244 ATCCAAATCAGATTTTGTCAAGTAGATGAGGAT-----ACCGGTATTAAATGTAATA 294
Qy 111 SerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyGlyPheGluGlyValAspPro 130
Db 295 CTAAATCAGGTCA-----GAAACAGAAATCAATGCACCCGGACCT 336
Qy 131 GlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIle----- 144
Db 337 AAAGTCACACACGCGCAACTTTCAACAACTTTTATCTCAAAATACGTCGTACTACAACGAC 396
Qy 145 GlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSer-----ValGlnGln 162
Db 397 GATATTGTTATTGAGCTGGAGGTACCGAACAGATATTCCAAGTGATGCTTATGCACAA 456
Qy 163 MetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAsp 182

Db 457 ATAGCACAAATCACTGAAAAAACTGGTCGCAACTAGTTGTTGATCCAGAAAAAGATCTT 516
Qy 183 PheGluArg-----TyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPhe 199
Db 517 GTTGAACAGATATTACCAATATCGA---CCATTATTATTAAACCAACAACAGATGAATTA 573
Qy 200 GluAlaValValGly---LysCysLysThrGluGluGluIleValGluArgGlyMetLys 218
Db 574 GAAGTAATGTTTAAACACTACAGTCAGAGATGATGAAGATGTTATAAAATATGTTAAAGAA 633
Qy 219 LeuIleAlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSer 238
Db 634 ATTTTGAAGGAGGGGCAATCAGTCATCATTTCTGTTGGCGATGGTGCAATATAT 693
Qy 239 LeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAla-----Gln 254
Db 694 GTTGATCA-----CATCAAGCATTAAGCTGTGAATCCACAAGGA 735
Qy 255 GluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeu 274
Db 736 CATGTAGTTAATACAGTAGGATCTGCTAGTAGTACAGTGGCAGGTATGTAGCAGGGTTG 795
Qy 275 AlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyValVal 294
Db 796 TCCATGGGTCTTAATATATAGATGAAGCT-----TTTCAACAGGCGCTAGCTTCA 843
Qy 295 ValGlyLysLeuGlyThrSerThr 302
Db 844 -----GGAACAGCGGACT 855

Search completed: November 26, 2002, 22:41:51.
Job time : 111 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 26, 2002, 22:04:38 ; Search time 113 Seconds
(without alignments)
1625.664 Million cell updates/sec

Title: us-09-912-020-325

Perfect score: 2393

Sequence: 1 MKYTLPEFERAGVMVGVDM.....FEDGCTTIKKIQDQKKG 477

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USFO.spool/US09912020/runat_25112002_091430_24365/app_query.fasta_1.647
-DB=PublishedApplications_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09912020@cgn.1.1.47 @runat_25112002_091430_24365
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Published Applications NA: *
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	2393	100.0	1434	10	US-09-741-669-276	Sequence 276, App
2	2393	100.0	1434	10	US-09-912-020-165	Sequence 165, App
3	2393	100.0	1434	10	US-09-815-242-6204	Sequence 6204, Ap
4	2265	94.7	1434	10	US-09-815-242-9661	Sequence 9661, Ap

5	1350.5	56.4	1425	10	US-09-815-242-7981	Sequence 7981, Ap	
6	197	8.2	927	10	US-09-815-242-7757	Sequence 7757, Ap	
7	191.5	8.0	918	10	US-09-815-242-6795	Sequence 6795, Ap	
8	190.5	8.0	1283	10	US-09-070-927A-532	Sequence 532, App	
9	182.5	7.6	6948	10	US-09-070-927A-4	Sequence 4, Appli	
C	10	181.5	7.6	880	10	US-09-974-300-747	Sequence 747, App
11	181.5	7.6	888	10	US-09-974-300-839	Sequence 839, App	
12	178	7.4	921	10	US-09-815-242-6938	Sequence 6938, Ap	
13	169	7.1	930	10	US-09-741-669-286	Sequence 286, App	
14	169	7.1	930	10	US-09-815-242-6309	Sequence 6309, Ap	
15	167	7.0	933	10	US-09-815-242-9878	Sequence 9878, Ap	
16	165	6.9	930	10	US-09-815-242-9785	Sequence 9785, Ap	
17	158	6.6	930	10	US-09-815-242-6083	Sequence 6083, Ap	
18	146	6.1	383	10	US-09-974-300-6200	Sequence 6200, Ap	
19	145	6.1	396	10	US-09-815-242-4344	Sequence 4344, Ap	
20	145	6.1	399	10	US-09-815-242-8205	Sequence 8205, Ap	
21	145	6.1	399	10	US-09-815-242-8777	Sequence 8777, Ap	
22	145	6.1	399	10	US-09-815-242-9053	Sequence 9053, Ap	
23	139.5	5.8	915	10	US-09-815-242-8919	Sequence 8919, Ap	
24	139.5	5.8	915	10	US-09-815-242-9055	Sequence 9055, Ap	
25	138.5	5.8	2565	10	US-09-815-242-4137	Sequence 4137, Ap	
C	26	134.5	5.6	1244	10	US-09-770-445-33	Sequence 33, Appl
27	131	5.5	4732	10	US-09-070-927A-62	Sequence 62, Appl	
28	129.5	5.4	942	10	US-09-815-242-6927	Sequence 6927, Ap	
29	128	5.3	362	10	US-09-878-574-347	Sequence 347, App	
C	30	124	5.2	396	10	US-09-815-242-3705	Sequence 3705, Ap
C	31	123.5	5.2	2350	10	US-09-070-927A-109	Sequence 109, App
32	122	5.1	11471	10	US-09-954-314-16	Sequence 16, Appl	
33	121.5	5.1	8145	10	US-09-070-927A-386	Sequence 386, App	
34	120.5	5.0	3535	10	US-09-220-091-16	Sequence 16, Appl	
C	35	120	5.0	272	10	US-09-741-669-89	Sequence 89, Appl
36	116.5	4.9	780	10	US-09-974-300-5390	Sequence 5390, Ap	
C	37	116.5	4.9	3717	10	US-09-904-065-1	Sequence 1, Appli
C	38	116.5	4.9	3723	10	US-09-904-065-3	Sequence 3, Appli
39	115.5	4.8	405	10	US-09-815-242-6416	Sequence 6416, Ap	
40	115	4.8	4851	9	US-09-712-363-116	Sequence 116, App	
41	114.5	4.8	2223	10	US-09-815-242-6279	Sequence 6279, Ap	
C	42	112	4.7	292	10	US-09-917-800A-1264	Sequence 1264, Ap
C	43	112	4.7	7287	10	US-09-070-927A-210	Sequence 210, App
44	111.5	4.7	912	10	US-09-815-242-9242	Sequence 9242, Ap	
45	111.5	4.7	1809	10	US-09-880-107-3376	Sequence 3376, Ap	

ALIGNMENTS

RESULT 1
US-09-741-669-276
; Sequence 276, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741.669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
US-09-741-669-276
Alignment Scores: 1.74e-257 Length: 1434
Pred. No.: 1.74e-257

Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-912-020-325 (1-477) x US-09-741-669-276 (1-1434)

```
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTAAACGTCGCAGAGTTTGAACGTCAGAGTGTATGTTGGTGTATGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProVal 40
Db 61 CTGGATCGTTACTGTTACGCGCCACCAGTCGTATCTCGCCGGAAGCGCGTCCG 120
QY 41 ValLysValIasnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTGAATACCATCGAAGAACGTCCTCGGCGCGCGCTAACGTGGCGATGAATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCTTCTCTCGGTGCTTAATGACGCCCTTGTCGGTGTACGGGCATTCACGATGACGCGCC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAlaValLysCysAspPheValSerValProThr 100
Db 241 CGCGTGAAGTAACTCTGGCGCGACGTCACGTCGAATGCGACTTCGTTCTGTACCGACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnIleLeuLeuArgLeuAsp 120
Db 301 CATCCGACCATTAACCAATTAACGGGTACTTCCCGCAACCAACACAGCTGATCCGCTGGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAGTTTCGAAGGTGTGTATCCGACGCCGCTGCACAGCGGATTAATCAGCGG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTGATTCGCGCGCTGGTGTCTTCTGACTACGCCAAAGGTGCGTGGCAAGCGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGCAGATGATCAACTCGCGCGCTAAAGCGGTGTTCCCGGTGCTGTATGATCAACAAAGGT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGGCGCTACCGCGCGCTACCGTGTATACGCCGAATCTCTCGGAATTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTTGTCGTAATGTAAGACCGAAGAGAGATTTCTGAGCGCGCATGAACCTGATT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCGGATTAACCAACTCTCGGCTCTGTTAGTGACCGCTTCCCAACAGGAGTATCGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCAGCGAGGAAGTGTATGACGTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaGlyAsnSerLeu 280
Db 781 GTGCGGGCCACACGGTGTATGCGGTCTCGCGGCAACGCTGCGCGGGTAAATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAAGAAGCCCTCTCTTTGCGCAATGCGCGCGCTGCGGTGCGTGGTCCGCAACCTGGAACC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTTCGCGGATCGAGCTGGAAAAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
```

```
Db 961 GGCGTGTATGACCGAAGCAAGTGAAGCTGGCGTGAAGCGCAGCGCTAAACCTGCTGAA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAGTGTGTATGATCAACACGCTGCTTTTGACATCTCGACGCGCGGACGCTCTCTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAlaAsnSerAspAlaSerThr 380
Db 1081 GCAATATGCCCGCAAGCTGGGTGACCCCTTGATGTTGCGGTCAACAGCGATGCTCCACC 1140
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGCTGAAAGGGATTCCTCCGCGGTAACCCACTCGAACAGCGTATGATGTTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCGACTGGAAGCGGTGACGCTGGGTAGTGTCTGTTTGAAGAGGACACGCGCGACGCTTG 1260
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATCGCGGGATCTTCCAGATCTGCTGGTGAAGCGCGGACTATAAACAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCGCGGAGTAAAGAGTCTGGGCAACGCTGGGAAAGTGTGCTCAACTTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLysGly 477
Db 1381 GPTTGTGCGACGACCAACATCATCAAGAGATCCCAACAGGATAAAAAAGGC 1431
```

RESULT 2

US-09-912-020-165

; Sequence 165, Application US/09912020

; Patent No. US20020045592A1

; GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Trauwick, John

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Froelich, Jamie M.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

; FILE OF INVENTION: ESCHERICHIA COLI

; FILE REFERENCE: ELITRA.001DV1

; CURRENT APPLICATION NUMBER: US/09/912,020

; PRIOR FILING DATE: 2001-07-23

; PRIOR APPLICATION NUMBER: 09/492,709

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/117,405

; PRIOR FILING DATE: 1999-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 165

; LENGTH: 1434

; TYPE: DNA

; ORGANISM: E. Coli

US-09-912-020-165

Alignment Scores:

Pred. No.: 1,74e-257 Length: 1434

Score: 2393.00 Matches: 477

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 10 Gaps: 0

US-09-912-020-325 (1-477) x US-09-912-020-165 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20

Db 1 ATCAAGTAACGCTGCAGAGTTTGAACGTGCAGGAGTGATGCTGTTGGTGTGATGTGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTGTACTGGTACGGCCCGCCACAGTCGTATCTCCCGGAAGCGCGGTGCCCGTG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTGAATACCATCAAGACGTCGCGGCGCGGCTAACTGGCGGATGATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArg 80
Db 181 GCTTCTCTCGGTGCTAATCACGCCCTTGGTTCGGGCTTGACGGCATTCACGATGCAGCGCG 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGAGTAATCTCTGCGCCAGCTCAACGTCATCGACTTCGTTCTGTACCGAGCG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATACCAAAATACGGGTACTTTCGCCCAACCAACAGCTGATCGCTGTGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAAGTTTCAAGGTGTTCATCCGACGCCCTGCAGAGCGGATTAATCAGGCG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTTCGATTGGCGCGTGGTGTCTTCTGACTACGCCCAAGAGTGGCTGGCAACGGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProTyrGly 180
Db 481 CAGCAGATGATCCAACATGCGCGCTAAAGCGGTGTTCCGGTGTGATGATCCAAAAGCT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGACGCTTACCGCGCGCTACGCTGTTAAACCCGAATCTCTCGAAATTTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTGTGCGTAATGTAAAGCCGGAAGAGAGATTGTTGAGCGCGCATGAACATGAT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCCGATTACGAACCTCTCGGCTCTGTTAGTGACCCGTTCCGGAACAGGGTATGCGTGTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCCAAGCCGAGGAAGTGTATGACGTATCC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGCGACACGTTGATGGCGTCTTGGCGGCAACCGCTGGCAGCGGGTAATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
Db 841 GAAGAAGCCTGCTCTTTTCCCAATGGCGCGCTGGCGTGGTGGTGGCGCAAACTGGGAACC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTTCCCGCATCAGCTGGAAAAATGCTGTACGTGGAGCTGCATACAGCGCTTT 960
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTGATGACCGAAGAGAACTGAAGCTGGCGGTAGCGGCGGCGGTAAACGTGGTGAA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAGTGGTATGACCAACAGGTGCTTTTGATCTCTCACCGCGCGGACGCTCTCTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAAAATGCCGCAAGGTGGGTGACCGCTTGTGATTGTTGGCGGTCAACAGCGATGCTCCAC 1140

QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGCTGAAAGGGGATTCGCCCGGTAACCCACTCGAACAGCGTATGATTGTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCACTGGAAGCGTGCAGCTGGGTAGTGTCTGTTTGAAGAGGACACGCGCAGCGCTTG 1260
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATCGCGGGATCTTGCAGATCTGCTGGTGAAGGCGGCGACTATAAACACAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCGCGAGTAAGAAGCTGGGCCAACGGTGGCGAAGTGTGTGCTCAACTTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnAspLysLysGly 477
Db 1381 GGTGCTGACGACCAACATCATCAAGAAGATCCAAACAGGATAAAAAAGC 1431

RESULT 3

US-09-815-242-6204
; Sequence 6204, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6204
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
US-09-815-242-6204

Alignment Scores:
Pred. No.: 1,74e-257 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-09-912-020-325 (1-477) x US-09-815-242-6204 (1-1434)

Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAAGTAAAGCTGCCAGAGTTTGAACGTGCAGAGGTGATGGTGTGATGATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTTACTGGTACGGGCCCCACCAGTGTATCTCGCCGGAAGCCGGTCCCCGTG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTGAATACCATCGAAGACGCTCCGGGGCGCGGTAACTGCGCGCATGATATC 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCTTCTCTCGGTGCTAATGTCAGCCTGTGTGCGGTGACGGGCATTGACGATGCAGCGGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 CGCTGTAGTAATCTCTGGCCGACGTCAACGTCAATCGACTTCGTTTCTGTACCGGACG 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATTTACCAATTTACGGGTACTTTCCCGCAACCAACAGCTGATCCGTCGTGAT 360
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAAGAAGGTTTCGAAGGTGTGATCCGCGAGCGCTGCACGAGCGGATTAATCAGCGC 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTTCGATTTGGCGCGCTGTGCTTTCTGACTACGCCAACGAGGTGCGCTGGCAAGCGTA 480
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGCAGATGATCCAACTCGCGCGGTAAACGGGTGTTCGGTGTGATGATCCAAAGGT 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACGCCGAATCTCTCGGAATTTGAA 600
Qy 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTGTGCGTAAATGTAAGACCGAAGAGAGATGTTGTAGCGCGCGCATGAAACTGATT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCCGATTACGAACCTCGGCTCTGTAGTGACCCGTTCCGACACAGGTATGTCGCTGCTG 720
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCAAAGCGCAGGAAGTGTATGACGTTACC 780
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGCGACACGGTGTATGGCGTCTCGCGCAACGCTGGCAGCGGGTAATTCGCTG 840
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAAGAGGCTGCTTCTTTGCCAATGCGCGCGCTGCGGTGGTGGTGGCGCAAACTGGGAACC 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyIleArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTTCGCGCATCGAGCTGGAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTGATGACGAAGAGAACTGAAGCTGGCGGTAGCGGCGCGGTAAACGGTGGTAA 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAGTGGTGTGACCAAGGTTCTTTGACATCTCTGCACGCGCGGCAGCTCTCTTATCTG 1080

Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAATGCCCGCAAGCTGGTGACCGCTTGATGTTGCCGTCAACAGCGATGCCCTCCACC 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnAlaMetIleValLeu 400
Db 1141 AAACGGCTGAAAGGGGATTCGCCCGCGGTAAACCCACATCGACACGCTATGATTGTGCTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCAGCTGGAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATCGCGGGATCTTCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCCGGGAGTAAAGAGTCTGGGCCAACCGTGGCAAGTGTGGTGTCTCAACTTTGAAGAC 1380
Qy 461 GlyCysSerThrThrAsnIleLeuLysLysIleGlnGlnAspLysLysGly 477
Db 1381 GGTGCTCGACACCAACATCATCAAGAAGATCCCAACAGGATAAAAAAGGC 1431
RESULT 4
US-09-815-242-9661
; Sequence 9661, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9661
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1434)
US-09-815-242-9661
Alignment Scores:
Pred. No.: 3,23e-243 Length: 1434
Score: 2265.00 Matches: 445
Percent Similarity: 97.48% Conservative: 19
Best Local Similarity: 93.49% Mismatches: 12
Query Match: 94.65% Indels: 0

DB: 10 Gaps: 0
US-09-912-020-325 (1-477) x US-09-815-242-9661 (1-1434)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
DB 1 ATGAAGTAAATCTGCCAGCGTTGAACGTGCGAGCGCTCATGTTGTGGGTGATGTAATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 61 CTGGATCGGTATGTGATGGCCCACTTCCCGTATTTACCGGAAGCGCGGTGCCCGG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
DB 121 GTTAAGGTAATACCGTTGAGGAACGTCCGGCGCGCGGCGAAGCGTGGCGATGAACATT 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
DB 181 GCGTCTCTGGGAGCGAACCCTGCTGGTCCGCCCTGACGGGTATTGATGACCGCGCGC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
DB 241 GCGCTGAGCAAAACGCTGCGGAGGTCAATGTGAAGTGCACCTTCGTTCTGTGCCGAGC 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnLeuIleArgLeuAsp 120
DB 301 CATCCGACGATTACCAAACTGCGCGTACTATCACGTAATACGACGCTCATTCGTCATTGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
DB 361 TTGAAGAAAGGCTTTGAGGCGGTGGACCGCGACCGCTGCATGAGCGGTATCAACCGCGC 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
DB 421 CTGGGATCATCGCGCGCTGTGTTGTCGATTATGCCAAAGCGCTCTGACCACCGTG 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
DB 481 CAGACTATGATTCCCTAGCGCGCCAGCGCGCGTCCGCGTGCATCATCGTAAAGGA 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
DB 541 ACGGATTTTGAACGTTTACCGCGCGCTGCTGCTGACGCCAACCTTCTGAAATTCGAG 600
QY 201 AlaValAlaGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
DB 601 GCGGTTCCGGGAAATGTAAACGGAAGACGAACTGGTTGAACCGCGCATGAAGCTCAT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
DB 661 GCGGATTACGACTTTCCGCGCTGTTGGTTCACGCGTTCCGAAACAGGGAATGACGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValThrAspValThr 260
DB 721 CAACCAATTAAGCGCGCTACATATGCGGACGCGAGCGCGAGGAGTTTATGATGTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
DB 781 GTGCGGGCGATACGCTGATCGCGTGTGGCGCGGACGCTGGCGCGGGAATACCCCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
DB 841 GAGGAGCGGTATTATTCGCCAATGCGCGCGCGGCGGTAGTGGTAGTAACCTCGGAGC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
DB 901 TCAACGGTTTCCCTATTGAGCTGGAACACGAGTGCAGCGCGCGCATACCGGCTTC 960
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaAlaArgLysArgGlyGlu 340
DB 961 GCGCTCATGCCAAGAGGAGTTGACAGCGCGCTGCCAGCGCGGTAAAGCTGCGGAG 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360

DB 1021 AAAGTGGTTCATGACCAACGCGCTTTTCGATATTTCTGACGCGGCCACGCTCTTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
DB 1081 GCGAACGCGCGCAAACTGGCGGACCGCTGATGTTGGCGGTCAATAGTACGCGCTCGACT 1140
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
DB 1141 AAACGCTCAAGAGCGAAAGCGCTCGGTTAATCCGCTCGAACAGCGTATGCTGCTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
DB 1201 GCGCGCTCGAGTGGTGCAGTGGGTGCTCTCTTTGAAGAGGATACGCGCAACGACTG 1260
QY 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTyrLysProGluGluIle 440
DB 1261 ATTGCGCGTATTCTGCGGATCTGCTGGTAAAGCGCGGACTATTAAGCGGGAAGATC 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
DB 1321 GCGGCGAGGAGAGGTCTGGCCACGCGCGGAGTTCATGCTGCTGAACCTTCGAAGAT 1380
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLys 476
DB 1381 GGTGTTGTCACGACCAATATCATCAAAAAGATCCAGACGCGAGCGAG 1428

RESULT 5

US-09-815-242-7981
; Sequence 7981, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlbeck, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7981
; LENGTH: 1425
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1425)
US-09-815-242-7981

Alignment Scores: 2.77e-141 Length: 1425
Pred. No.: 1350.50 Matches: 274
Score:

Percent Similarity:	72.73%	Conservative:	70
Best Local Similarity:	57.93%	Mismatches:	128
Query Match:	56.44%	Indels:	1
DB:	10	Gaps:	1

US-09-912-020-325 (1-477) x US-09-815-242-7981 (1-1425)

QY	1	MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet	20
Db	1	ATGAAGTTGTCCATGCCCGCTTTCACACAGSCCCGGTGTTGGTGGTGCCTGATGTGATG	60
QY	21	LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal	40
Db	61	CTCAGACCCTACTGTCATGCGCGCACTTCGGCGCATTTTCGCGGAGGCCCGGTGCCGGT	120
QY	41	VallYsValAsnThrILeGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle	60
Db	121	GTCCCGCTCGAACACGACGAGAGCGCCCGCGCGCGCCGCCAACGTCGCGCTGAACATC	180
QY	61	AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg	80
Db	181	GC CGCGCTGGCGCGCAGGCCCTTCCTGCTGGGGTCAACGCGCGACGAGCGCGCCGAC	240
QY	81	AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr	100
Db	241	AGCTTGGCCAACACGCCCTCAAAGCGCGTGGAGTGSACGCGCGCTTCCACGCGCATCGATAGC	300
QY	101	HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp	120
Db	301	CAGCGGACCATCTGCAAGCTGCGGGTCTATGAGTGCACCAGCAACTGCTGGCGGTCGAC	360
QY	121	PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla	140
Db	361	TTCGAGGAACGTTCCGC---ACGACGCGCGCGCCCTGGCCGTGCAGCTCGAGTCGCTG	417
QY	141	LeuSerSerILeGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal	160
Db	418	CTGCCCAAGGTCAAGTGTCTGGTGTCTCGACCTACGGCAAGSGCGCGGTACAGAACCAC	477
QY	161	GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly	180
Db	478	CAGGTGTGTATCCAGCGCGCGCGCGCGCACATTCCGGTACTGCCCGATCCCAGGCG	537
QY	181	ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu	200
Db	538	AAGGACTTCGCCATCTATCTCGCGCGCGACGCTGATCACCACCACTGTCGGAATTCGAG	597
QY	201	AlaValValGlyLysCysLysThrGluGluGluIleValIcLuArgGlyMetIlysLeuIle	220
Db	598	ACCATCTCGCGCTTGC GCGCGACGAAGCGCAACTGGTGC CAAGGGCCAGCGCTGATG	657
QY	221	AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu	240
Db	658	AGCGAACTCGACCTGCTGTCTGGTGTGACCCGCGCGACATGGCATGACCTGCTGCT	717
QY	241	GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr	260
Db	718	CGGCATGGCGCAGCGCCCTCGACCTGCGCGCGCGCGCGGCGGAAAGTGTTCGACCTCAC	777
QY	261	GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu	280
Db	778	GGTCCCGCGCATACGGTCACTCCACCTTGGCGCGCGCGCTTTCGCGCGCGGAGGAGCTG	837
QY	281	GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValIcLyLysLeuGlyThr	300
Db	838	CCTTCGCGGTGGTCTGCGCAACCTGGCGCGCGCGCATGTGGTTCGGCAAGCTGGGTACC	897
QY	301	SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe	320
Db	898	CGCGCATCAGCGCGCCCACTGCGTCTCGCGGGTGCAGCGCGACAGGGTTCGAGCGT	957
QY	321	GlyValMetThrCluGluLeuLysLeuAlaAlaAlaAlaAlaArgGlyArgGlyGlu	340

Alignment Scores:

```
Pred. No.: 5,67e-13 Length: 927
Score: 197.00 Matches: 86
Percent Similarity: 41.85% Conservative: 50
Best Local Similarity: 26.46% Mismatches: 153
Query Match: 8.23% Indels: 36
DB: 10 Gaps: 11

US-09-912-020-325 (1-477) x US-09-815-242-7757 (1-927)

Qy 10 ArgAlaGlyValMetValGlyAspValMetLeuAspArgTyrTrpTyrGlyProThr 29
Db 4 CAACGAAAGTACTGGTGGGAGCCTGAACATGAGCTGGAGTGGCGCGCGCGC 63
Qy 30 -----SerArgIleSerProGluAlaProValProValValValAsn 44
Db 64 CTGCCACGGCGCGGAGACCTCGCGGGCAGTCTTCATCACCGTT----- 111
Qy 45 ThrIleGluGluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSerLeu 63
Db 112 -----CCCGCGGCAAGGAGGCAACCAAGCGGTCGCGCGCGCGCTCTC 156
Qy 64 GlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSer 83
Db 157 GCGCGGAGGTGGCGATGCTGGCTTGGCGAGCATGCTACGGCGGACAGTTGTAC 216
Qy 84 LysSerLeuAlaAspValAsnValLysCysAspPheValSer---ValProThrHisPro 102
Db 217 CGTGGCTACAGCGCGGAGGCTGACTGCCAGGGCGTCTGAGCGGGTGGCGGGAGTCC 276
Qy 103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGlu 122
Db 277 AGCGGAGTGGCGCTGATCGTGGTGGACGACGACGACGACGCGTCTGTCGTCGCC 336
Qy 123 GluGlyPheGluGlyValAspProGlnProLeu-----HisGluArgIleAsnGlnAla 140
Db 337 GCGCGCAACGCGCCACTGTCCCGCGCGTGTGGCGGCCACGACACTTCTCTGGAGCAG 396
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 397 -----GCGCAGGTGGTAGTCTGCCAATGGAAGCCCGCTGGAGCGGTT 441
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspPro----- 178
Db 442 GGCATGTGCTGCGCGCGGCCCATCGCTGGCGAAGCGGTGATCTCAACCGCGCGCG 501
Qy 179 -----LysGlyThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeu 196
Db 502 GCCACCGCGAGTGGCGCGCGGAGTGGCTGCGGTGGTGGTACTACCTGTCCTCCCAACGAG 561
Qy 197 SerGluPheGluAlaValValGlyLysCysLysThrGluGluGluIleValGlu----- 214
Db 562 ACCGAAACGAACTGCTC-----TGCGCGCTGCGGTGGTGGTCTCTGGAGAGCGCC 612
Qy 215 ---ArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuLeuValThrArgSer 233
Db 613 GGGCGTGGCGCGCGGCGCTCGCGAATGGCGCGCGCGGAGTGTGTCACCTCGCG 672
Qy 234 GluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAla 253
Db 673 GCCAGGGCGCCCTGCTGGTAGGCGAGGGCGGGTTCGAG---CATTTCCCGGTGGCGCGG 729
Qy 254 GlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThr 273
Db 730 GTCAGGGCGTGGATACACGCGCGCGCGGACACCTTCGTCGGCGGTTCGCGCGGCC 789
Qy 274 LeuAlaAlaGlyAsnSerLeuGluAlaCysPhePheAlaAsnAlaAlaAlaGlyVal 293
Db 790 CTTGCGCGCGCGCTCGAGGAGCGCGCGCATCGCTTCGCGCGAGCGCGCGCGCGATC 849
Qy 294 ValValGlyLysLeuGlyThrSerThrValSerProIle-----GluLeuGluAsnAla 311
Db 850 TCGGTGACCCCGCTCGCGCGCGACACCTCGATCCCTCGCGCGGAGGAGTGAAGTGGCGGCG 909
```

```
Qy 312 ValArgGlyArgAla 316
Db 910 CTCGTTGGCGAGGCT 924

RESULT 7
US-09-815-242-6795
; Sequence 6795, Application US/09815242
; Patent NO. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6795
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(918)
US-09-815-242-6795

Alignment Scores:
Pred. No.: 2.29e-12 Length: 918
Score: 191.50 Matches: 72
Percent Similarity: 42.72% Conservative: 63
Best Local Similarity: 22.78% Mismatches: 142
Query Match: 8.00% Indels: 39
DB: 10 Gaps: 11

US-09-912-020-325 (1-477) x US-09-815-242-6795 (1-918)

Qy 13 ValMetValValGlyAspValMetLeuAspArgTyrTrpTyrGlyProThrSerArgIle 32
Db 16 GTTACAATTATTCGGAAGTATCAATTTAGAC-----ACAACATTAAAGATC 60
Qy 33 SerProGluAlaProValProValValLysValAsnThrIleGluGluArgProGlyGly 52
Db 61 AAA---GAATATGCCAAAACAGCGCGGAAACAATTACGCCATTGAAATTATTTACAGCGGT 117
Qy 53 -----AlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeu 69
Db 118 GGTGCTAAAGGAGCAACCAAGCAGCTCGCAGCGAAGCGTTTCAGCGCGCAACATATTTT 177
Qy 70 ValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeuAla----- 87
Db 178 ATTGTGTGCTGGGAATGATGATGCTGCGCTGAGCTATGATGACTGATGATTAATGAGTCAAGT 237
```

QY 88 AspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLysLeu 107
Db 238 GAAATTAATTA-----ACTGGGTCCACCATTTA 267
QY 108 ArgValLeuSerArgAsnGlnLeuIleArgLeuAsp----- 120
Db 268 GAGAAAACAGCTACTGGCCAAGCCTTATTATGGTAGATACAGCTGGTGAATAAGCAATT 327
QY 121 -----PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsn 138
Db 328 ATGATTATCCGGGAGCCCAATAACGCCTTTACGCCAAAACAAGTACAGGAA-----CAT 381
QY 139 GlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAla 158
Db 382 CAAGAAATTAAGCAAAAGTGAATTTTGTGATGCGCAATTT---CAAGAGCGGATGAT 438
QY 159 SerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspPro 178
Db 439 AGTACGATTGCAGCGTTTAAATTTGCTAAAGAGCAGCGCTCAAAACGATTTTAAATCCC 498
QY 179 -----LysGlyThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrPro 194
Db 499 GCACCTGCGTTAGAACAAAGTCTCTGAAGAAATTAATAACGCTGACAGATATGATTGTACCA 558
QY 195 AsnLeuSerGluPheGluAlaValValGlyLysCysLysThrGluGluGluIleValGlu 214
Db 559 AATGAACAGAACCAACCAATTTTAACAGGCATTAATAACAGATGACGAGTATGCGT 618
QY 215 ArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuValThrArgSerGlu 234
Db 619 AAAGCGGCGACAGCACTTCAATTAATAGGATTAAGAGCAGATTAATTAACAGTAGTAGT 678
QY 235 GlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGln 254
Db 679 AAAGAGACCTTTTATGACGTCATGACGAAGTGGTATT---GTGCTGCTTTTAAAGTG 735
QY 255 GluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeu 274
Db 736 AAAGCGGTTGATACAAACGCGTGTGCGGATCTTTTATTGGCGCATTAAGTAGATATATTA 795
QY 275 AlaAlaGly---AsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyVal 293
Db 796 GAAAAAGATTTTACGATTTTGAAGAAGCTATTCGTATGGAACAAGCGCTCTCGTTG 855
QY 294 ValValGlyLysLeuGlyThrSerThrValSerProIleGluLeuGlu 309
Db 856 ACTGTTCAACGTTTGGAGCCCAACCTTCGATTCCTTATCAACACGAA 903

RESULT 8

US-09-070-927A-532
; Sequence 532, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 962
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 532:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 532:
US-09-070-927A-532

Alignment Scores:
Pred. No.: 4-8e-12 Length: 1283
Score: 190.50 Matches: 81
Percent Similarity: 40.48% Conservatives: 53
Best Local Similarity: 24.47% Mismatches: 126
Query Match: 7.96% Indels: 71
DB: 10 Gaps: 11
US-09-912-020-325 (1-477) x US-09-070-927A-532 (1-1283)
QY 13 ValMetValValGlyAspValMetLeuAspArgTyrTrpTyr----- 26
Db 230 ATGATTGTAGTGGTAGTTTAAACGTTGATTAATTATCTGTATCAACAAATTTACCACAT 289
QY 27 ---GlyProThrSerArgIleSerProGluAlaProValProValValLysValAsnThr 45
Db 290 AACGGCAAAACAATTTTGTGCTAGCTAGCTAAATTTCCCT----- 331
QY 46 IleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAla 65
Db 332 -----GGCGAAAAAGGATTAAATCAAGCTGTGGCTTAAACAAACTGGGCCAT 379
QY 66 AsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSer 85
Db 380 CAGGCCACCTTAATTGGTTGTTTAGGCTCAGACACAGATCGGAATTACTTTATATAAGAA 439
QY 86 LeuAlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThr 105
Db 440 TTAGAAAAATACCACTGTTAGCAGAT-----GGTATCACA 475
QY 106 LysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp----- 120
Db 476 CGTATTCAAGATACACTGAGCTGTCAAGCTTATATTATGTTGAGACACGCGGACTCC 535
QY 121 -----PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArg 136
Db 536 ATGATTTCATCTTACTCTGGTGCACACACAGCGCTTACGCTTAAATAAATTCGCACAGCAA 595
QY 137 IleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAla 156
Db 596 AAACACCTATTTTATGATGCCAGTTTTGTCTCATTCAAAACAGAAAT-----CCT 646
QY 157 LeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIle 176
Db 647 TTGCTGCTGTAGAGAAAGCTTGTGAATTCGCCCAACATTCAGGGGTACCGATATTATTA 706
QY 177 AspPro-----LysGlyThrAspPheGluArg 185
Db 177 AspPro-----LysGlyThrAspPheGluArg 185

```
Db 707 AACCCAGCTGCTATTCATCATATTCAGTGGAATATACAGAAAAGGTGATTTT----- 760
Qy 186 TyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLys 205
Db 761 -----TTTCATTCCTCCCAATCAAGATGAATATTGGAACTTCAACCCAGAT 802
Qy 206 CysLysThrGluGluGlu-----IleValGluArgGlyMetLysLeuIleAla 221
Db 803 ACTGGTACATTAGAAGAGAGCGGCTACTTCTTACAAATGGCGGTAAAAATGTC--- 859
Qy 222 AspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGln 241
Db 860 -----ATTGTTACTTAGGNAAGAAAGCGTG---TTATTAAAA 895
Qy 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGly 261
Db 896 ACACCCCAAGTGTGCCACTATTTCCCTGCACAGAAAATATAGCTGTGGATAGCACTGGT 955
Qy 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 956 GCCAGTCACAGCTTATCAGCGGCTCGCTCCTCTATCTTTCGAAAGGTATTCGCCACTGAA 1015
Qy 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr--- 300
Db 1016 GCAGCCATTCAAATAGCAATTCAGGCTGCAGGATTTTCAGTTTCTTAAAGAGGGGTGATT 1075
Qy 301 ---SerThrValSerProIleGluLeuGluAsn 310
Db 1076 GATTCTTGTAGTCACGTCACCTTTGGAAAAAT 1108

RESULT 9
US-09-070-927A-4/c
: Sequence 4, Application US/09070927A
: Patent No. US20020120116A1
: GENERAL INFORMATION:
: APPLICANT: Charles A. Kunsch
: Patrick J. Dillon
: Steven Barash
: TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
: NUMBER OF SEQUENCES: 982
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/070,927A
: FILING DATE: 04-May-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/046,655
: FILING DATE: 1997-05-16
: APPLICATION NUMBER: 60/044,031
: FILING DATE: 1997-05-06
: APPLICATION NUMBER: 60/066,009
: FILING DATE: 1997-11-14
: ATTORNEY/AGENT INFORMATION:
: NAME: Kenley K. Hoover
: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6948 base pairs
```

```
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-070-927A-4

Alignment Scores:
Pred. No.: 4.25e-10 Length: 6948
Score: 182.50 Matches: 77
Percent Similarity: 41.09% Conservative: 66
Best Local Similarity: 22.13% Mismatches: 158
Query Match: 7.63% Indels: 47
DB: 10 Gaps: 12

US-09-912-020-325 (1-477) x US-09-070-927A-4 (1-6948)
Qy 29 ThrSerArgIleSerProGluAlaProValProValValLysValAsnThrIleGluGlu 48
Db 6936 ACATTAAAGAGTCAAA---GAAATGCCAAACACGCGGCAAAACAATTCACGCCATTGAACAT 6880
Qy 49 ArgProGlyGly-----AlaAlaAsnValAlaAlaMetAsnIleAlaSerLeuGlyAla 65
Db 6879 TTTACACAGCGTGTGTGTAAAGGAGCAACCAACAGCTTCGCGGAAACGTTTCAGGCCA 6820
Qy 66 AsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSer 85
Db 6819 GAAACATATTTTATTGTCGTGGGAAATGATGCGCTGGAGCTATGATCACTGATTTA 6760
Qy 86 LeuAla-----AspValAsnValLysCysAspPheValSerValProThrHisProThr 103
Db 6759 ATGAGTCAAGATGAATTAATTTA-----ACTGGG 6730
Qy 104 IleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp----- 120
Db 6729 GTCACCACTTTAGAGAAAACAGCTACTGCGCAACGCTTTATRTGTGTAGATAACGCTGGT 6670
Qy 121 -----PheGluGluGlyPheGluGlyValAlaAspProGlnProLeuHis 134
Db 6669 GAAATAGCAATTATGATTACGCGAGGAGCAATAACGCTTTACGCCAAAAACAAGTCCAG 6610
Qy 135 GluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLys 154
Db 6609 GAA-----CATCAAGAAATTTATTGAAAAAAGTAGTATTGTGATTGCCCAATTT---GAA 6559
Qy 155 GlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProVal 174
Db 6558 AGTCGGATTGATAGTACGATTGCGAGCGTTTAAATTTGCTAAAAAGCAGGCGTCAAAACG 6499
Qy 175 LeuIleAspPro-----LysGlyThrAspPheGluArgPyrArgGlyAlaThr 190
Db 6498 ATTTTAAATCCGCACTGCGTCTAGAACCAAGTTCCTGAAGAATTTACTAAACGTGACAGAT 6439
Qy 191 LeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrGluGlu 210
Db 6438 ATGATTGTACCAATTAAGAACAGAACCGCAATTTTAAACAGGCATTAAATAACACAGATGAA 6379
Qy 211 GluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuVal 230
Db 6378 GCGAGTATGCGTAAAGCGCGAGAGCACTTCATCAATTAGGRTTGAACAGTAGTAATATT 6319
Qy 231 ThrArgSerGluGlnGlyMetSerLeuGlnProGlyLysAlaProLeuHisMetPro 250
Db 6318 ACAGTAGTAGTAAAGGCGCTTTTATGACGCTCAATGACGCAAGTGGTATT---GTGCGCT 6262
Qy 251 ThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeu 270
Db 6261 GCTTTTAAAGTGAAGCGGTGTGATACCAACGCGCTCTGCGGATACTTTTATTGGCGCATTA 6202
Qy 271 AlaAlaThrLeuAlaAlaGly---AsnSerLeuGluGluAlaCysPhePheAlaAsnAla 289
Db 6201 AGTAGTATATGAAAAAGATTTTAGCAATTTTGAAGAAGCTATTCGTTATGGAAACAG 6142
Qy 290 AlaAlaGlyValValValGlyLysLeuGlyThrSerThrValSerProIle-----Glu 307
```



```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6309
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(930)
US-09-815-242-6309

Alignment Scores:
Pred. No.: 7,53e-10 Length: 930
Score: 169.00 Matches: 77
Percent Similarity: 39.49% Conservative: 47
Best Local Similarity: 24.52% Mismatches: 154
Query Match: 7.06% Indels: 36
DB: 10 Gaps: 9

US-09-912-020-325 (1-477) x US-09-815-242-6309 (1-930)

QY 9 GluArgAlaGlyValMetValVal---GlyAspValMetLeuAspArgTyrTrpTyrGly 27
   :: ||||| ::||| ||| ::|
Db 4 CAAACGCGAGCGCTCGTGTCTTGGCAGCATTAATGCTGACCCACATTT-----54

QY 28 ProThrSerArgIleSerProGluAlaProValProValValValValAsnThrIleGlu 47
   :: ||| ||| ||| ||| |||
Db 55 ---CTTAATCTCAATCTTTTCTACTACGCGGAAACCGTAACCGGTAAACCATATCAG 111

QY 48 GluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn 66
   ||||| ||||| ||||| ||||| |||||
Db 112 GTTGCAATTTGGCGGCAAGCGCGAATCAGCGTGTGGCTGCTGGCGGTAGCGGTGCGAAT 171

QY 67 AlaArgLeuValGlyLeuThrGlyIleAlaAspAlaAlaArgAlaLeuSerLysSerLeu 86
   :: ||||| ||| ::| ::| ::|
Db 172 ATCGCGTTATTGCTTACGGGTGATGACAGCATTTGGTGAGACGCTTCCCGACGACTC 231

QY 87 AlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLys 106
   ||| ||| ::| ::| ::|
Db 232 GCCACTGATAACATTT-----GATATTACTCG 258

QY 107 LeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGlyPheGlu 126
   :: ||||| ::| ||| ||| |||
Db 259 GTCAGCGTGATCAAGCGCAATCAACAGAGGTGGCGCTGATTTTGTAAATGCG---GAA 315

QY 127 GlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSer-----142
   ||| ::| ::| ::| ||| ||| |||||
Db 316 GGTGAGAAATCATCGGTATTCAGCGCGGCTAAATGCTGCCCTTTCCCGCGCGCTGGTG 375

QY 143 -----SerIleGlyAlaLeuValLeuSerAspTyrAlaLys 154
   :: ||||| ::| ::| ::|
Db 376 GAAGCGCAACGCTAGCGTATTGCCAGCGCTCAGCATTTATTATGCGAG-----CTGGAA 429

QY 155 GlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProVal 174
   ||| ||||| ::| ::| ::|

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 26, 2002, 20:59:44 ; Search time 3602 Seconds
(without alignments)
3329.530 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFRAGVMVGVDM.....FEDGCTNIKKIQDKKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlp
-Q=/cgn2_1/USFTO_spool/US09912020/runat_25112002_091429_24305/app_query.fasta_1.647
-DB=Pending_Patents_NA_Main -QMT=fastap -SUFFIX=p2n.rnmp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09912020 -CGN1_1_3142 -runat_25112002_091429_24305
-NCPU=6 -ICPU=3 -NO_XLPX -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending_Patents_NA_Main: *
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq: *
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq: *
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq: *
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq: *
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq: *
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq: *
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq: *
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq: *
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq: *
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq: *
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq: *
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq: *
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq: *
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq: *
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq: *
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq: *
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq: *
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq: *
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq: *
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq: *
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq: *
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq: *
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq: *
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq: *
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq: *
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq: *
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq: *
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq: *

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq: *
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq: *
31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq: *
32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq: *
33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq: *
34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq: *
35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq: *
36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq: *
37: /cgn2_6/ptodata/2/pna/US099D_COMB.seq: *
38: /cgn2_6/ptodata/2/pna/US100A_COMB.seq: *
39: /cgn2_6/ptodata/2/pna/US100B_COMB.seq: *
40: /cgn2_6/ptodata/2/pna/US101A_COMB.seq: *
41: /cgn2_6/ptodata/2/pna/US101B_COMB.seq: *
42: /cgn2_6/ptodata/2/pna/US102A_COMB.seq: *
43: /cgn2_6/ptodata/2/pna/US102B_COMB.seq: *
44: /cgn2_6/ptodata/2/pna/US6000_COMB.seq: *
45: /cgn2_6/ptodata/2/pna/US6001_COMB.seq: *
46: /cgn2_6/ptodata/2/pna/US6002_COMB.seq: *
47: /cgn2_6/ptodata/2/pna/US6003_COMB.seq: *
48: /cgn2_6/ptodata/2/pna/US6004_COMB.seq: *
49: /cgn2_6/ptodata/2/pna/US6005_COMB.seq: *
50: /cgn2_6/ptodata/2/pna/US6006_COMB.seq: *
51: /cgn2_6/ptodata/2/pna/US6007_COMB.seq: *
52: /cgn2_6/ptodata/2/pna/US6008_COMB.seq: *
53: /cgn2_6/ptodata/2/pna/US6009_COMB.seq: *
54: /cgn2_6/ptodata/2/pna/US6010_COMB.seq: *
55: /cgn2_6/ptodata/2/pna/US6011_COMB.seq: *
56: /cgn2_6/ptodata/2/pna/US6012_COMB.seq: *
57: /cgn2_6/ptodata/2/pna/US6013_COMB.seq: *
58: /cgn2_6/ptodata/2/pna/US6014_COMB.seq: *
59: /cgn2_6/ptodata/2/pna/US6015_COMB.seq: *
60: /cgn2_6/ptodata/2/pna/US6016_COMB.seq: *
61: /cgn2_6/ptodata/2/pna/US6017_COMB.seq: *
62: /cgn2_6/ptodata/2/pna/US6018_COMB.seq: *
63: /cgn2_6/ptodata/2/pna/US6019_COMB.seq: *
64: /cgn2_6/ptodata/2/pna/US6020_COMB.seq: *
65: /cgn2_6/ptodata/2/pna/US6021_COMB.seq: *
66: /cgn2_6/ptodata/2/pna/US6022_COMB.seq: *
67: /cgn2_6/ptodata/2/pna/US6023_COMB.seq: *
68: /cgn2_6/ptodata/2/pna/US6024_COMB.seq: *
69: /cgn2_6/ptodata/2/pna/US6025_COMB.seq: *
70: /cgn2_6/ptodata/2/pna/US6026_COMB.seq: *
71: /cgn2_6/ptodata/2/pna/US6027_COMB.seq: *
72: /cgn2_6/ptodata/2/pna/US6028_COMB.seq: *
73: /cgn2_6/ptodata/2/pna/US6029_COMB.seq: *
74: /cgn2_6/ptodata/2/pna/US6030_COMB.seq: *
75: /cgn2_6/ptodata/2/pna/US6031_COMB.seq: *
76: /cgn2_6/ptodata/2/pna/US6032_COMB.seq: *
77: /cgn2_6/ptodata/2/pna/US6033_COMB.seq: *
78: /cgn2_6/ptodata/2/pna/US6034_COMB.seq: *
79: /cgn2_6/ptodata/2/pna/US6035_COMB.seq: *
80: /cgn2_6/ptodata/2/pna/US6036_COMB.seq: *
81: /cgn2_6/ptodata/2/pna/US6037_COMB.seq: *
82: /cgn2_6/ptodata/2/pna/US6038_COMB.seq: *
83: /cgn2_6/ptodata/2/pna/US6039_COMB.seq: *
84: /cgn2_6/ptodata/2/pna/US6040_COMB.seq: *
85: /cgn2_6/ptodata/2/pna/US6041_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2393	100.0	1434	1	PCT-US02-03987-6204
2	2393	100.0	1434	18	US-09-492-709A-165
3	2393	100.0	1434	29	US-09-741-669-276
4	2393	100.0	1434	31	US-09-815-242-6204
5	2393	100.0	1434	34	US-09-912-020-165
6	2393	100.0	1434	39	US-10-072-851-6204
					Sequence 6204, App
					Sequence 165, App
					Sequence 276, App
					Sequence 6204, App
					Sequence 165, App
					Sequence 6204, App

7 2299 96.1 1449 16 US-09-252-691-3421 Sequence 3421, Ap
8 2299 96.1 1449 16 US-09-252-691C-3421 Sequence 3421, Ap
9 2265 94.7 1434 1 PCT-US02-03987-9661 Sequence 9661, Ap
10 2265 94.7 1434 31 US-09-815-242-9661 Sequence 9661, Ap
11 2265 94.7 1434 39 US-10-072-851-9661 Sequence 9661, Ap
12 2227 93.1 1500 18 US-09-489-039A-2037 Sequence 2037, Ap
13 2007 83.9 1476 21 US-09-543-681A-2462 Sequence 2462, Ap
14 1968 82.2 27693 33 US-09-897-516-1462 Sequence 1461, Ap
15 1968 82.2 27693 33 US-09-897-516-1462 Sequence 1462, Ap
16 1968 82.2 27693 33 US-09-897-516-1462 Sequence 1463, Ap
17 1968 82.2 27693 33 US-09-897-516-1462 Sequence 1464, Ap
18 1968 82.2 27693 33 US-09-897-516-1465 Sequence 1465, Ap
19 1968 82.2 27693 33 US-09-897-516-1466 Sequence 1466, Ap
20 1968 82.2 27693 33 US-09-897-516-1467 Sequence 1467, Ap
21 1968 82.2 27693 33 US-09-897-516-1468 Sequence 1468, Ap
22 1968 82.2 27693 33 US-09-897-516-1469 Sequence 1469, Ap
23 1968 82.2 27693 33 US-09-897-516-1470 Sequence 1470, Ap
24 1968 82.2 27693 33 US-09-897-516-1471 Sequence 1471, Ap
25 1968 82.2 27693 33 US-09-897-516-1472 Sequence 1472, Ap
26 1968 82.2 27693 33 US-09-897-516-1473 Sequence 1473, Ap
27 1968 82.2 27693 33 US-09-897-516-3225 Sequence 3225, Ap
28 1968 82.2 27693 33 US-09-897-516-3226 Sequence 3226, Ap
29 1968 82.2 27693 33 US-09-897-516-3227 Sequence 3227, Ap
30 1968 82.2 27693 33 US-09-897-516-3228 Sequence 3228, Ap
31 1968 82.2 27693 33 US-09-897-516-3229 Sequence 3229, Ap
32 1968 82.2 27693 33 US-09-897-516-3230 Sequence 3230, Ap
33 1968 82.2 27693 33 US-09-897-516-3231 Sequence 3231, Ap
34 1968 82.2 27693 33 US-09-897-516-3232 Sequence 3232, Ap
35 1968 82.2 27693 33 US-09-897-516-3233 Sequence 3233, Ap
36 1968 82.2 27693 33 US-09-897-516-3234 Sequence 3234, Ap
37 1968 82.2 27693 33 US-09-897-516-3235 Sequence 3235, Ap
38 1968 82.2 27693 33 US-09-897-516-3236 Sequence 3236, Ap
39 1968 82.2 27693 65 US-60-215-161-1461 Sequence 1461, Ap
40 1968 82.2 27693 65 US-60-215-161-1462 Sequence 1462, Ap
41 1968 82.2 27693 65 US-60-215-161-1463 Sequence 1463, Ap
42 1968 82.2 27693 65 US-60-215-161-1464 Sequence 1464, Ap
43 1968 82.2 27693 65 US-60-215-161-1465 Sequence 1465, Ap
44 1968 82.2 27693 65 US-60-215-161-1466 Sequence 1466, Ap
c 45 1968 82.2 27693 65 US-60-215-161-1467 Sequence 1467, Ap

ALIGNMENTS

RESULT 1
PCT-US02-03987-6204
; Sequence 6204, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits
; FILE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA-028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6204
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
PCT-US02-03987-6204
Alignment Scores:
Pred. No.: 1,09e-204 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-912-020-325 (1-477) x PCT-US02-03987-6204 (1-1434)
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTAAACGTCGCCAGATTGAACGTGCAGGAGTATGTTGGTGTGATGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTTACTGGTACGGCCCCACAGTCGTATCTCGCCGCAACGCCGGTGCCTG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTAAATACCATCGAAGAACCTCCGGCGCGGCTAACCTGCGCATGAATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCTTCTCTCGTCTAATGCACGCTGGTGGGTTCACGGGCATTCACGATGACGCGCGC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGAGTAAATCTCTGGCGGACGTCAACGTCAAAATGGACTTCGTTCTGTACCGAGC 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATTAACAAATACGGGTACTTCCCGCAACCAACAGCTGATCCCTCTGGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTGAGAGAGTTTCAAGGTGTGTATCCGCACGCCCTGCACGAGCGGATTAATACGGCG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTCGATTGGCGCGCTGCTGCTTCTGACTAGCCCAAGGTGCGTGGCAAGCGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuLeuLeuProLysGly 180
Db 481 CAGCAGATGATCAACTGGCGGTAAAGCGGGTGTTCGGTGTGATTGATCCAAAGGT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGAGCGCTACCGCGCGCTACGCTGTTAACGCCGAATCTCTCGGAATTTGA 600
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTTGTCGTAATGTAAGACCGAAGAGAGATTGTTGAGCGCGGCATGAACATGATT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGlnGlnGlyMetSerLeuLeu 240
Db 661 GCGATTACGAATCTCGGCTCTGTTAGTACCCCTTCGGAACAGGGTATGTCGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCCAAGCCGAGGAGTGTATGACGTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGGACACGCTGATTGGCTGCTCGCGGCAACGCTGCGCAGCGGTAATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAAGAAGCCTGCTTCTTTCGCAATGCGCGGCTGGCTGGTTCGCGCAACCTGGGAACC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACGGTTTCGCGGATTCGAGCTGGAAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTGATGACCGAAGGAACTGAAGCTGGCGGTAGCGGCACGCGCTAAACCTGGTGAA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1020

Db 1021 AAAGTGGTATGACCAACGGTCTCTTTCACATCTGCGCGGCGACGCTCTTATCTG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyValSerLeuValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAATATGCCGCAAGCTGGTGACCGCTGATGTTGGCTGCAACAGCGATGCCCTCCACC 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGGCTGAAGGGATTCGCCGCCGGTAAACCCACATCGAACAGCGATGATGTCGTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCACTGGAAGCGTGCAGTGGTAGTGTCTGTTGAAGAGGACACGCGCAGCGCTTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATCGCGGGATCTTCCAGATCTGCTGGTGAAGGCGGCGATATAAACCAAGAGATTT 1320
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCCGGAGTAAAGAGTCTGGGCCAACGGTGGCGAAGTGTGCTCAACTTTGAAGAC 1380
Qy 461 GlyCysSerThrAsnIleIleLysLysIleGlnGlnAspLysGly 477
Db 1381 GGTGCTCGACGACCAACATCATCAAGAAGATCCCAACAGGATAAAAAAGGC 1431

RESULT 2

US-09-492-709A-165

; Sequence 165, Application US/09492709A

; GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Trawick, John

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Froelich, Jamie M.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

; FILE REFERENCE: ELITRA.001A

; CURRENT APPLICATION NUMBER: US/09/492,709A

; CURRENT FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 165

; LENGTH: 1434

; TYPE: DNA

; ORGANISM: E. Coli

US-09-492-709A-165

Alignment Scores:

Pred. No.:	1.09e-204	Length:	1434
Score:	2393.00	Matches:	477
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	18	Gaps:	0

US-09-912-020-325 (1-477) x US-09-492-709A-165 (1-1434)

Qy 1 MetLysValThrProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTAAACCTGCCAGAGTTTGAACGTGCAGAGTGATGTTGGTGTATGTGATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTTACTGTGTACGCCGCCACCACTGATATCTCGCCGGAAGCGCGCGCTG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTGAATACCATCGAAGAACGTCCGGCGCGCGGTAAACGTGCGGATGATATC 180

Db 1261 ATCCGGGATCTTGCAGATCTGCTGGTGAAGGCGGCGACTATAAACAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCCGGGAGTAAGAATCTGGGCCAACGGTGGCGAAGTGTGTGCTCACTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleLysLysIleGlnGlnAspLysLysGly 477
Db 1381 GGTGTCTCGACGACCAACATCATCAAGAGATCCACAGAGATAAAGGC 1431

RESULT 3

US-09-741-669-276
; Sequence 276, Application US/09741669
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
US-09-741-669-276

Alignment Scores:

Pred. No.: 1,09e-204 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29 Indels: 0
DB: Gaps: 0

US-09-912-020-325 (1-477) x US-09-741-669-276 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAAGTAAACGCTGCAGAGTTTGAACGTCGAGGAGTGATGTTGGTGATGTGATG 60
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGTTTACTTGGTACGCGCCACAGTCGTATCTCGCGGAAGCGCGGTGCCGCTG 120
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAAGTGAATACCATCGAAGAACGTCGCGGCGCGCGCTAACGTCGGATGATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCCTCTCTCGTCTAATGCACGCCCTGGTGGGTTTACGGGCATTCGACGATGCGCGCGC 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCCTGAGTAAATCTCTGGCGGACGTCACGTCAAATGGCACTTCGTTTCTGTACCGACG 300
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACCATTAACAAATTACGGGTACTTTCCCGCAACCAACACGATGATCCGCTGGAT 360
QY 121 PheGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTTGAAGAGGTTTCAAGAGTGTGTATCCGACGCCCTGCACGAGGGGATTAATCAGGCG 420

RESULT 4

US-09-815-242-6204
; Sequence 6204, Application US/09815242
; GENERAL INFORMATION:

QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGAGTTCGATTGGCGCGCTGGTGGTCTTCTGACTACGCCAAAGGTGCGCTGGCAAGCETA 480
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGCAGATGATCAACTGGCGCTAAAGCGGTGTTCGCGTGTGATTGATCAACAAAGGT 540
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACCGATTTTGAGCGCTACCGCGCGCTACGCTGTTAACGCCGAATCTCTCGGAATTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCTGTTGTCGTAATGTAAGACCGAAGAGAGATGTTTGAGCGCGCATGAAATGATT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCCGATTACGAACTCGGCTCTGTTAGTGACCCGTTCCGAACAGGATATGTCGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCCAGCGCAGGAAGTGTATGACGTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGCGACACGCTGATTGGCGTCTGCGCGCAACGCTGGCAGCGGTATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr 300
Db 841 GAAGAAAGCCTGCTTCTTGGCAATGCGCGCTGCGGTGCTGCGCAACATGGGAACC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCCACCGTTTCGCGCATCGAGCTGGAAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
QY 321 GlyValMetThrGluGluGluLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGCTGATGACCAAGAGGAACTGAAGCTGGCGTAGCGCGCAGCGGTAAACGTTGGTGA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAAGTGTGATGACCAACGCTGCTTTGACATCTGCACCGCGGCGACGCTCTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCAATGCCCGCAAGCTGGGTGACCGCTTGATTGTTGCCGCTCAACGCGATGCTCCACC 1140
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGCTGAAGGGGATTCGCCCGCGGTAAACCCACTCCGAACAGCGTATGATGTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCGACTGGAAGCGGTGCGACTGGTGTGCTGCTGTTTGAAGAGGACACGCCGCGCTG 1260
QY 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyAspTyrLysProGluGluIle 440
Db 1261 ATCCCGGGGATCTTGCAGATCTGCTGGTGAAGGCGGCGACTATAAACAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCCGGGAGTAAAGAGTCTGGGCCCAACGGTGGCGAAGTGTGTGCTCACTTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleLysLysIleGlnGlnAspLysLysGly 477
Db 1381 GGTGTCTCGACGACCAACATCATCAAGAGATCCACAGAGATAAAGGC 1431

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 6204
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
; US-09-815-242-6204

```

Alignment Scores:		
Pred. No.:	1.09e-204	Length:
Score:	2399.00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	100.00%	Mismatches:
Query Match:	100.00%	Indels:
DB:	31	Gaps:
		0
US-09-912-020-325 (1-477)	x US-09-815-242-6204 (1-1434)	

Qy	1	MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet	20
Db	1	ATGAAAGTAACGCTGCAGAGTTGAACGTGCAGGAGTGTGGTGGTGTGATGTGATG	60
Qy	21	LeuAspArgTyrTrpPyrGlyProThrSerArgIleSerProGluAlaProValProVal	40
Db	61	CTGGATCGTTTACGGACGGCCACCATCGTATCTCGCCGAAAGCGCGGTGCCCGG	120
Qy	41	ValLysValAsnThrIleGluGluArgProGlyGlyValAlaAlaAsnValAlaMetAsnIle	60
Db	121	GTTAAGTGAATACCATCGAAGAACCTCCGGCGGCGCGCTAACGTGGCGATGAATATC	180
Qy	61	AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArg	80
Db	181	GCATTCTCTCGGTCTAATGCACGCTTGGTCGGGTTCACGGCATTCACGATGCAGCGCG	240
Qy	81	AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr	100
Db	241	GCCTCAGTAAATCTCTGGCCGACGTCAACGTCAATGCGACTTCCTGTCTGACCGACG	300
Qy	101	HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp	120
Db	301	CATCCGACCATTAACAAATACGGGTACTTCCCGCAACCAACAGCTGATCGCTCUGGAT	360
Qy	121	PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla	140

Db	361	TTTTGAAGAGGGTTTCGAAGGGTGTGATCCGACGCGCTGCACGAGCGGATTAATTCAGCGC	420
Qy	141	LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal	160
Db	421	CTGAGTTCGATTCGGCGCGTGGTCTTCTGACTACGCCAAGAGTGCCTGGCAAGCGTA	480
Qy	161	GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly	180
Db	481	CAGCAGATATCCAACTGCGCGCTAAAGCGGTGTTCGGTCTGCTGATTCACCAAAAGGT	540
Qy	181	ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu	200
Db	541	ACCGATTTCAGCGCTACCCGCGCGCTACGCTGTAAACGCCGAATCTCTCGGAATTTCAA	600
Qy	201	AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle	220
Db	601	GCCTGTCTCGTAAATGTAAAGCCGAGAGAGATGTGTAGCCGCGCATCAAACTGATT	660
Qy	221	AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu	240
Db	661	GCGGATTACGAATCTCGGCTCTGTAGTAGACCCGTTCGCAACAGGGTATGCTCGCTGCTG	720
Qy	241	GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr	260
Db	721	CAACCGGGTAAAGCGCGCTGCATATGCCAACCCAAAGCGCAGGAAGTGTATGACGTTACC	780
Qy	261	GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu	280
Db	781	GGTCGGCGCACAGCTGATTGGCTCTCGCGGCACACGCTGGCAGCGGGTAATTCGCTG	840
Qy	281	GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThr	300
Db	841	GAAGAAGCCCTGCTTTTCCCAATCGCGCGCTGGCGTGGTTCGCGCAAACTCGGAACC	900
Qy	301	SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe	320
Db	901	TCCACGGTTTCGCGCGATCGAGCTGGAAAATGCTGTACGTGGACGTGCAGATACAGGCTTT	960
Qy	321	GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu	340
Db	961	GGCGTGATGCCAAGAGGAATGAAGCTGGCGGTAGCGGCAGCGCGTAACGTTGGTCAA	1020
Qy	341	LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu	360
Db	1021	AAAGTGTGTATGACCAACCGTGTCTTTGATCTCTGCACGCGCGGCACGTCTCTATCTG	1080
Qy	361	AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr	380
Db	1081	GCAATGCCCGCAAGCTGGTGACCGCTTGATGTTCGGCTCACGCGCATGCCTCCACCC	1140
Qy	381	LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu	400
Db	1141	AAACGGGTGAAGGGGATTCGCCCGCGGTAAACCCACTCGAACACGCGTATGATTGTGCTG	1200
Qy	401	GlyAlaLeuIleAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu	420
Db	1201	GGCGCACTGGAACGGTTCGACTGGGTAGTGTCTGTTGAAGAGGACACCGCCGACGGCTG	1260
Qy	421	IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle	440
Db	1261	ATCGCCGGGATCTTGCAGATCTGCTGTGTGAAGGGCGGCATATAAACAGAGAGAGATT	1320
Qy	441	AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp	460
Db	1321	GCCGGGAGTAAGAAGTCTGGGCCAACGCTGGCGGAAGTGTGGTGCTCAACTTTGAAGAC	1380
Qy	461	GlyCysSerThrThrAsnIleLysLysIleGlnAlaAspLysLysGly	477
Db	1381	GGTTGCTCGACGACCAACATCATCAAGAAGATCCCAACAGGATAAAAAAGGC	1431

; Sequence 165, Application US/09912020
; GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

; FILE OF INVENTION: ESCHERICHIA COLI

; FILE REFERENCE: ELITRA.001DV1

; CURRENT APPLICATION NUMBER: US/09/912,020

; CURRENT FILING DATE: 2001-07-23

; PRIOR APPLICATION NUMBER: 09/492,709

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/117,405

; PRIOR FILING DATE: 1999-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 165

; LENGTH: 1434

; TYPE: DNA

; ORGANISM: E. Coli

US-09-912-020-165

Alignment Scores:

Pred. No.: 1,09e-204 Length: 1434
Score: 2393.00 Matches: 477
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 34 Gaps: 0

US-09-912-020-325 (1-477) x US-09-912-020-165 (1-1434)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
DB 1 ATGAAGTAACGTCGCCAGAGTTTGAACGTGCAGGAGTGTGTGTGTGTGTGTGTGTGT 60
QY 21 LeuAspArgTrpTrpGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 61 CTGGATCGTTACTGGTACGCGCCACACAGTCGTATCTCGCGGAGCGCGCGTCCCGTG 120
QY 41 ValLysValAsnThrIleGluAtrgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
DB 121 GTTAAGTGAATACCATCGAAGAACGTCCGCGCGCGCGCGGTACGTGGCGATGATATC 180
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValIleLeuThrGlyIleAspAspAlaAlaArg 80
DB 181 GCITCTCTCGGTGCTAATGACGCTGTCGGTGTGACGGGTCATTGACGATGACGCGCG 240
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
DB 241 CGCTGAGTAAATCTCTGGCGCGAGTCACAGTCAAAATCGCACTTCGTTTCTGTACCG 300
QY 101 HisProThrIleThrLysLeuAtrgValLeuSerArgAsnGlnGlnLeuIleAtrgLeu 120
DB 301 CATCCGACCAATACCAATACCGGTACTTCCCGCAACCAACAGCTGTATCCGTCGTGGAT 360
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluAtrgIleAsnGlnAla 140
DB 361 TTGGAAGAAGTTTCGAAGGTGTGTATCCGCGCGCGTGCACGCGCGGATTAATACGGCG 420
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTrpAlalysGlyAlaLeuAlaSerVal 160
DB 421 CTGAGTTCGATGGCGCGCTGGTCTTCTGACTACGCCAAGGTGGCGTGGCAAGCGTA 480
QY 161 GlnGlnMetIleGlnLeuAlaAtrgAlaGlyValProValLeuIleAspProLysGly 180
DB 481 CAGCAGATGATCCAACTGGCGCGGTAAAGGGGTGTTCCTGGTGTGTATGATCCAAAGGT 540

QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
DB 541 ACCGATTTTGGCGCTACCGCGCGCTACGCTGTAAACGCGAATCTCTCGGAATTTGAA 600
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
DB 601 GCTGTTGTCGGTAAATGTAAGACCGAAGAGATGTTGAGCGCGCATGAACACTGATT 660
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
DB 661 GCGGATACCAACTCTCGGCTCTGTAGTACCGCTCCGAACAGGATGTGCTGCTGCTG 720
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
DB 721 CAACCGGGTAAAGCGCGCTGCATATGCCAACCAACGCGCAGGAAGTGTATGAGTTTACC 780
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
DB 781 GGTGCGGGCGACACGCTGATTGGCTCTCTGGCGCAACGCTGGCAGCGGTAATTCGCTG 840
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
DB 841 GAAGAAGCCTGCTTCTTTGCCAATGCGCGCTGGCGTGGTGGTGGCAACACTGGGAACC 900
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
DB 901 TCCACGGTTTCCGCGATCGAGCTGGAATGCTGTACGTGGACGTGCAGATACAGGCTTT 960
QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaAlaArgLysArgGlyGlu 340
DB 961 GCGCTGATGACCGAAGAGGAACCTGAAGCTGGCGGTAGCGCGCAGCGCGTAACCTGGTGA 1020
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
DB 1021 AAAGTGTGTGATGACCAACGCTGCTTTGCATCTCTTGACCGCGGCGACGCTCTTATCTG 1080
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
DB 1081 GCAATGCCCGCAAGCTGGGTGACCGCTTGATGTTGTCCTCAACACGCGATGCCCTCCACC 1140
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
DB 1141 AAACGGCTGAAAGGGGATCCCGCGCGGTAAACCCATCGAACAGCGCTATGATTGCTG 1200
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
DB 1201 GCGCGACTGGAAGCGGTGCGACTGGGTAGTCTGCTTTGAAGGACACGCGCGCGCTTG 1260
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
DB 1261 ATCCCGGGATCTGCCAGATCTGCTGGTGAAGGCGCGGACTATAAACCCAGAGAGATT 1320
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
DB 1321 GCGCGGAGTAAAGAGTCTGGCGCAACGCTGGCGAAGTGTGGTGGCTCAACTTTGAAGAC 1380
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnAspLysLysGly 477
DB 1381 GGTGCTCGACGACCAACATCATCAAGAGATCCACAGGATATAAAGGC 1431

RESULT 6

US-10-072-851-6204

; Sequence 6204, Application US/10072851

; GENERAL INFORMATION:

; APPLICANT: Carr, Grant J.

; APPLICANT: Xu, H. Howard

; APPLICANT: Foulkes, J. Gordon

; APPLICANT: Zamudio, Carlos

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

Alignment Scores:

```
Pred. No.: 3 1e-196 Length: 1449
Score: 2299.00 Matches: 452
Percent Similarity: 98.95% Conservative: 19
Best Local Similarity: 94.96% Mismatches: 5
Query Match: 96.07% Indels: 0
DB: 16 Gaps: 0

US-09-912-020-325 (1-477) x US-09-252-691-3421 (1-1449)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
DB 19 ATGAAAGTAACACTGCCAGATTGAACGTGTGGAGTTATGTTGTCGGTGTGATG 78
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 79 CTGATCGCTACTGTGATGGCGGACGACCGCATCTCCGGAAGCACCGGTACCGGTG 138
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
DB 139 GTAAGGTGCACACCATTTGAAGCGTCCCGGTGGCGGCAACCGTGGCGATGAACATA 198
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
DB 199 GCCTCTCTGGGCGACATTCGGCTCTGGCTGACCGGCATTGACGATGCGGCGGT 258
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
DB 259 GCCTGAGCAAGTCGTGGCGGACGTGAATGTGAAGTGTGACTTCGTTCTGCTCGGACC 318
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
DB 319 CACCCCAACCATCACCAAGCTTCGCGTCTGCGGTAAACGACGAGCTAATCCGCTCGAC 378
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
DB 379 TTTTGAAGAGGGTTTGAAGCGTTGATCCCGAGCCCTGCACGAGCGCATTAACCAAGGG 438
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
DB 439 TTGGGCAATATTGGCGCGTGTGCTGTGATATGCCAAGCGCGCTGGCGGACGGT 498
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
DB 499 CAGACCATGATCAGCTGGCGGCAAGCGTTCGCGTGTGTGATCGACCCGGAAGGC 558
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
DB 559 ACCGATTTTGAGCGCTATCGCGCGCAACCCCTGCTGACGCCAACCTCTCTGAATTCGAA 618
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
DB 619 GCGGTGGCGGCAAGTGCACAACTGAAGAGAGCTGTTGAGCGCGCATGAAATCATC 678
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGlnGlnGlyMetSerLeuLeu 240
DB 679 GCTGATTTTCGAGTGTCTGCGCTGTGTTGACCCGTTCCGAGCAGGCGATGACGCTGCTG 738
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValThrAspValThr 260
DB 739 CAACCGGGCAAGCGCGCTGTCATATGCCAACCCGAGCGGAGGAAGTACGACGTGAC 798
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
DB 799 GGTGGCGGTGATACGCTGATGGCGTGTGGCGGCGACGCTGGCGGGGTAACTCCCTG 858
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
DB 859 GAAGAAGCATGCTACTCTCGCGAAGCGCGGCGGCGGTGCTGCTCGTAACTACCGGAC 918
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
DB 919 TCCACCGTTTCGCCAATTGAGCTGGAAACGCGGTGCGTGGCGGTGCGGATACCGGCTTT 978
```

```
QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
DB 979 GCGCTCATCAGCAGAGAGCTGAAGTGGCGTTGCCGCCGCCAAACCGGTGAA 1038
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
DB 1039 AAAGTGTGATGACGAACCGCGTGTTCGACATCTCATCGCGGCCACCGTGTCTATCTG 1098
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
DB 1099 GCGAATCGCGAAGCTGGCGACCGCTGATGTGGCGTCAACAGCATGCTCAACC 1158
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluAlaPheMetIleValLeu 400
DB 1159 AAACGTCTGAAGGTGAAACCGCTCCGGTTAACCCCTGGAGCAGCGCATGATAGTGCTC 1218
QY 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420
DB 1219 GCGCGCTGGAAGCGTGGACTGGGTGGTCTCATTTGAAGAAGACACACCCGCGACGCTG 1278
QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGluGluIle 440
DB 1279 ATTGCGGGAATTTGCTGACCTGCTGGTGAAGGGCGGTGATTACAAACCGACGAAATC 1338
QY 441 AlaGlySerLysGluValTyrAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
DB 1339 GCGAGTGAAGAGAGCTGGCGCAATGGTGTGAAGTAAATGGTGTCTCACTTTGAGGAC 1398
QY 461 GlyCysSerThrThrAsnIleLysLysIleGlnGlnAspLysLys 476
DB 1399 GGGTGTTCACCAACCAACATCATTAAGAAGATCCAGAAAGACAGTCAG 1446
```

RESULT 8

```
US-09-252-691C-3421
; Sequence 3421, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 3421
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-3421
```

```
Alignment Scores:
Pred. No.: 3 1e-196 Length: 1449
Score: 2299.00 Matches: 452
Percent Similarity: 98.95% Conservative: 19
Best Local Similarity: 94.96% Mismatches: 5
Query Match: 96.07% Indels: 0
DB: 16 Gaps: 0
```

US-09-912-020-325 (1-477) x US-09-252-691C-3421 (1-1449)

```
QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
DB 19 ATGAAAGTAACACTGCCAGATTGAACGTGTGGAGTTATGTTGTCGGTGTGATG 78
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
DB 79 CTGATCGCTACTGTGATGGCGGACGACCGCATCTCCGGAAGCACCGGTACCGGTG 138
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
```

Db 139 GTAAAGCTCGACACCATTTGAAGAGCGTCCCGTGGCGGCAAAAGTGGCGATGAACATA 198
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArg 80
Db 199 GCCTCTCTGGGCGCACATTCGCGTCTGGTGGCCCTGACCGCATTCAGCATGGCGCGT 258
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 259 GCGCTGAGCAAGCTCGTGGCGGACGTGAATGTGAAGTGTGACTTCGTTCTCTCCGACC 318
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 319 CACCAACCATCACCAAGCTTCGCGTGTGTCGCGTAACACGACGCTAATCCGCTCGAC 378
Qy 121 PheGluGluClyPheGluClyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 379 TTGTGAAGAGGTTTGAAGCGCTTATCCGAGCGCTGACAGCGCATTAACAGGCG 438
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 439 TTGGCAATATTGGCGCGTGTGCTGTCTGATTATGCCAAAGCGCGTGGCGAGCGTG 498
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 499 CAGACCATGATCCAGCTGGCGGCAAGCACGCTTCGCGTCTGATCGACCCGAAAGCG 558
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 559 ACCGATTGTGAGCGGTATCGCGGCGAACCTGCTGACGCCAAACCTCTCTGAATTCGAA 618
Qy 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 619 GCGGTGGCGGCACTGCAAACTGAAGAAGAGTGTGTGACGGCGCATGAATCATC 678
Qy 221 AlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGlnGlnMetSerLeuLeu 240
Db 679 GCTGATTTCGAGCTGTCTCGCTGTGTGTGACCCGTTCCGAGCAGGCGATGACGCTGTG 738
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 739 CAACCGGCAAGCGCGCTGCTATATGCCAACCCAGGCGCAGAAAGTGACACGTGACG 798
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 799 GGTCCGCTGATACGTGATTGCGTGTGTCGCGGCGACGCTGCGCGGGTAACTCCCTG 858
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 859 GAAGAAGCATGCTACTTCGCAACCGCGCGGCGGCGTGTGTCGCTAACTCCGAGCG 918
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 919 TCCACCGTTTCGCCAATTGAGCTGGAACACGCGGTGCGTGGCGGTGCGGATACCGGCTTT 978
Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 979 GCGCTCATGACCAACAGCAGCTGAAGTGGCGGTGCGCGCGCGCAACCGCGGTGAA 1038
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1039 AAAGTGGTGTATCAGCAACCGCGTGTGACATCCTGTGATGCGGCGCACGTGCTATCTG 1098
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1099 GCGAATGCGCGCAAGCTGGCGGACCGCTGATTGTGGCGGTCAACAGCGCATGCCCAACC 1158
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1159 AAACGCTGAAGGGTGAACCGCTCCGGTTAACCGCTGGGACGCGCATGATAGTGCTC 1218
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420

Db 1219 GCGCGCTCGAAGCGGTGGACTGGGTGTCTCATTTGAAGAAGACACCCCGAGCGCCTG 1278
Qy 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1279 ATTGCGCGNATTTTGCCTGACCTGCTGTGAAGGCGGTGATTACAAACAGAGCAATC 1338
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1339 GCAGGTAGCGAAGAGCTCGGCGAATGCTGTGAAGTAATGCTCAACTTTGAGGAC 1398
Qy 461 GlyCysSerThrAsnIleLeuLysIleGlnGlnAspLysLys 476
Db 1399 GGGTGTTCACCAACCATCATTAAGAAGATCCAGAAAGACAGTCAG 1446

RESULT 9
PCT-US02-03987-9661
; Sequence 9661, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elitra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9661
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
PCT-US02-03987-9661

Alignment Scores:
Pred. No.: 3,48e-193 Length: 1434
Score: 2265.00 Matches: 445
Percent Similarity: 97.48% Conservative: 19
Best Local Similarity: 93.49% Mismatches: 12
Query Match: 94.65% Indels: 0
DB: 1 Gaps: 0
US-09-912-020-325 (1-477) x PCT-US02-03987-9661 (1-1434)

Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAGTAATCTGCGACGCTTGAACGTGCAGGCGCTCATGTTGCTGCTGATGTAATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGAGTCGCTATTTGTTATGTCGCTTTCACCGGAGGCGGCGGCGGCGGCGGCGG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAGTGAATATACCGTTGAGGAACGTCGCGGCGGCGGCGGCGGCGGCGGCGG 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 GCGTCTCTGGAGCGAAGCGCGCTGCTGTCGCGCTGACGGGTATTGATCACGCCGCGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGGTGAGCAAAACGCTGGCGGAGTCAATGTGAAGTGGCGCTTCCTTCTGTGGCGAG 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACGATTACCAAACTGCGGCTACTATCAGCTAATCAGCAGCTCATCTCTGTAT 360
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140

```
Db 361 TTTGAGAGAGCGTTTGAGGGCGTGGAGCCGCGCGTTCATGAGCGTATCAACACGCG 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 421 CTGGATTCGATCGCGCGCGTGGTATTGTCGATATGCGAAGGCGCTCTGACACGCGTG 480
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGACTATGATTTCCCTAGCGCGCGAGCGCGCGTGCCTCATCGATCCGAAGGA 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACGGATTTTGAAGCTTACCGCGCGCGCACGCTGCTGACGCGCAACCTTTCTGAATTCGAG 600
Qy 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 CGGTTTCGGGGGAAATGTAACGGAAGACGAAGTGTGTAACCGCGCATGAAGCTCAT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCGGATTCAGACCTTTCCGCGCTGTTGGTCACGCGTTCCGACAGGGAATGACGCTGCTG 720
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnLaglnGluValTyrAspValThr 260
Db 721 CAACCGAATAAAGCGCGCTACATATGCCAGCAGCGCGAGGAAGTTTATGATGTATAC 780
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaGlyAsnSerLeu 280
Db 781 GGTGCGGGCGCATACGGTGTGCGGTGCTGGCGGCGAGCGTGGCGGCGGAATACCGCTG 840
Qy 281 GluGluAlaCysPheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAGGAGCGGTGTTATTTCCGCAATGCGCGCGCGGTAGTGTAGTAACTCGGGAGC 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCAACGGTTTCCCGCTATTGAGCTGGAAACGCGAGTGGCGGCGAGCGCGGATACCGGCTC 960
Qy 321 GlyValMetThrGluGluGluLysLeuAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGTCTACCGCAAGAGAGATTGAGACAGCGCTGCCAGCGCGTAAAGCGTGGCGAG 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAAGTGTCTATGACCAACGCGGTTTCGATATCTGCACGCGGCGCCACGCTCTTATCTG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCGAACCGCGCAACCTGGGCGAGCGGCTGATTGTTCCGGTCAATAGTACGCGCTCGACT 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAAGCTCTGAAGCGGAAACCGCTCCGGTTAATCCGGTCCGAACAGCGTATGATCGTCTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCGCTGAGTGGTGGTGGTGTCTCTTTTCAAGAGGATACGCGCGCAACGACTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGluGluIle 440
Db 1261 ATTTCCGGTATTCTGCGCGATCTGCTGGTAAAGCGCGGCGGACTATAGCGCGGAAGATC 1320
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCGGGCAGCAAGAGGTCTGGGCGCAACGCGCGGAAGTCAATGGTGTGAACCTCGAAGAT 1380
Qy 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspTyrLys 476
Db 1381 GGTGTTCACGACCAATATCATCAAAAAGATCCAGACCGGAGCGAG 1428
```

RESULT 10

US-09-815-242-9661

; Sequence 9661, Application US/09815242

```
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9661
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
; US-09-815-242-9661
```

Alignment Scores:

Pred. No.:	3 48e-193	Length:	1434
Score:	2265.00	Matches:	445
Percent Similarity:	97.48%	Conservative:	19
Best Local Similarity:	93.49%	Mismatches:	12
Query Match:	94.65%	Indels:	0
DB:	31	Gaps:	0

US-09-912-020-325 (1-477) x US-09-815-242-9661 (1-1434)

```
Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 1 ATGAAAGTAAATCTGCCAGCGTTTGAACGTGCAGCGTCATGGTTGGTGTGATGTAATG 60
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 61 CTGGATCGCTATTGGTATGGCCGACCTTGCCTGATTTTACCGGAAAGCGCGGTGCCGTG 120
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 121 GTTAAGGTAATACCGTTGAGGAACGTCCTGGGCGGCGCGGCGGCGGATGAACATT 180
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 181 CGGTCTCTGGGAGGAACGCCGCTCTGCTCGGCTGACGGGTATTGATGACGCGCGCGC 240
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 241 GCGCTGAGCAAAACGCTGGCGGAGGTCAATGTGAAGTGCAGCTTCGTTTCTGTGCCGACG 300
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 301 CATCCGACGATTACCAAACTGCGCGTACTATCAGCTAATCAGCAGCTCATTCGCTCTTGTAT 360
```

```
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 361 TTTCAAGAAAGGCTTTGAGGGCGTGGACCGCAGCGTTCGATGAGCTATCAACACGCG 420
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlalysGlyAlaLeuAlaSerVal 160
Db 421 CTGGATTCGATCGCGCGCTGGTATTCTCGGATTATGCCAAGGCGCTCGACCAAGCGTG 480
Qy 161 GlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 481 CAGACTATGATTTCCCTAGCGCGCAGCGCGCGTGCCTGCTATCGATCCGAAGAAGGA 540
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 541 ACGGATTTGAACGTTACCGCGCGCCACGCTGCTGAGCGCAACACCTTCTGATTCGAG 600
Qy 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 601 GCGGTTGCGGGGAATGTAAAGCGAAGACCACTGTTGAACGCGCATGAAGCTCATT 660
Qy 221 AlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 661 GCCGTTACGACCTTTCCGCGCTGTTGTCACGCGTTCCGACAGGAATGACGCTGCTG 720
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 721 CAACCGAATAAAGCGCGCTACATATGCCGACGCGCAGGAGGTTATGATGTTACC 780
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaGlyAsnSerLeu 280
Db 781 GGTGCGGCGGATACGGTATCGCGCTGCTGCGCGCGACGCTGCGCGGGGAAATACCGTG 840
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 841 GAGGAGCGGTATTTCGCCAATCGCGCGCGCGGTAGTGGTAGGTAACCTCGGACG 900
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 901 TCAACGGTTTCCCTATTAGCTGGAACACGACGTCGCGGACGCGGATACCGGCTTC 960
Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 961 GCGGTCATGACCGAAGAGGAGTTGAGACAGCGCGCTCGCGACGCGGTAAAGCGTGGGAG 1020
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1021 AAGTGGTCATGACCAACGCGGTTTTCGATATCTGACGCGGCGCCAGCTCTCTATCTG 1080
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1081 GCGAAGCGCGCAAACTGGGCGACCGCTGATTCTTGGGTCATAGTGACGCGCTCGACT 1140
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1141 AAACGTCGAAAGCGAAGCGGTCGCGGTTAATCCGTCGAAACAGCGTATGATCGTGGTG 1200
Qy 401 GlyAlaLeuGluAlaValAspTyrValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1201 GCGCGCTGGAGTCGCTGCGACTGGTGTCTCTTTTGAAGAGGATAGCGCCCAACGACTG 1260
Qy 421 IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1261 ATTCGCGTATTCGCGGATCTGCTGTAAAGCGCGGCTATAAGCGCGGAAGAGATC 1320
Qy 441 AlaGlySerLysGluValThrAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1321 GCGGCGACGGAAGAGGTCTGGCGCAACGCGCGCAAGTCAATGGTGTGCTCAACTTCGAAGAT 1380
Qy 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAspLysLys 476
Db 1381 GGTGTTCCACGACCAATATCATCAAAAGATCCAGACCGGAGCGGAG 1428
```

RESULT 11

US-10-072-851-9661

: Sequence 9661, Application US/10072851

: GENERAL INFORMATION:

: APPLICANT: Carr, Grant J.

: APPLICANT: Xu, H. Howard

: APPLICANT: Foulkes, J. Gordon

: APPLICANT: Zamudio, Carlos

: APPLICANT: Haselbeck, Robert

: APPLICANT: Ohlsen, Kari L.

: APPLICANT: Zyskind, Judith W.

: APPLICANT: Wall, Daniel

: APPLICANT: Trawick, John D.

: APPLICANT: Yamamoto, Robert T.

: APPLICANT: Roemer, Terry

: APPLICANT: Jiang, Bo

: APPLICANT: Boone, Charles

: APPLICANT: Bussey, Howard

: TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit

: FILE OF INVENTION: Proliferation

: FILE REFERENCE: ELITRA.028A

: CURRENT APPLICATION NUMBER: US/10/072,851

: PRIOR FILING DATE: 2002-02-08

: PRIOR FILING DATE: 2001-02-09

: NUMBER OF SEQ ID NOS: 15811

: SOFTWARE: FastSeq for Windows Version 4.0

: SEQ ID NO 9661

: LENGTH: 1434

: TYPE: DNA

: ORGANISM: Salmonella typhi

: FEATURE:

: NAME/KEY: CDS

: LOCATION: (1)...(1434)

US-10-072-851-9661

Alignment Scores:

Pred. No.: 3,48e-193 Length: 1434

Score: 2265.00 Matches: 445

Percent Similarity: 97.48% Conservative: 15

Best Local Similarity: 93.49% Mismatches: 12

Query Match: 94.65% Indels: 0

DB: 39 Gaps: 0

US-09-912-020-325 (1-477) x US-10-072-851-9661 (1-1434)

Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20

Db 1 ATGAAAGTAAATCTGCCAGCGCTTTGAACGTGCGCGCTCATGGTGTGGTGGTAAATG 60

Qy 21 LeuAspArgTyrTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40

Db 61 CTGATCGCTATTGGTATGGCCCTTCCTGCTATTTCCACCGAAGCGCGCGTGGTGGT 120

Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60

Db 121 GTTAAAGTAAATACCGTTGAGAACGCTCCGGCGCGCGGCGGACGTCGCGCATGAACATT 180

Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80

Db 181 GCGTCTCTGGGAGGAAACGCGCGCTGCTGCGCTGAGCGGTATGATGACGCGCGCGCG 240

Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100

Db 241 GCGCTGAGCAAAACGCTGGCGGAGGTCAATGTGAAGTCGACTTCGTTCTGTGCGCGAG 300

Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120

Db 301 CATCCGACGATTACCAAACTGCGCGTACTATACGTAATACGACGCTCATTCGCTTGTAT 360

Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140

Db 361 TTTGAAGAAGGCTTTGAGGGCGTGGACCGCGCGCGTGGTATGACGCTATCAACACGCG 420


```
Db 787 CAGCGCGGACGTCGCCCGCTGCATATGCGGACCCAGGCCAGGAAGTGTGACGTGACC 846
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 847 GCGCGCGCGACACCGTGTATGGCTACTGCGCGGACCCCTGGCGTCCGCGCAACACCTG 906
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaGlyValValIleGlyLysLeuGlyThr 300
Db 907 GAAGAGGCTGCTATTTCGCTAACCGCGCGCGGCGGTGGTGGTGGTGGTGGTGGTGGT 966
Qy 301 SerThrValSerProIleGlyLeuGluAlaValAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 967 TCCACCGTTCCCTGTGCGAGCTGAAAAACCGGTCCTCGCGCGCGGCTGAGACCGGCTT 1026
Qy 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 1027 GCGGTGTGTAGCGAAGAGAGGCTCAACAGCGCGGTGGCGGCGCGCGCAAGCGCGCGAG 1086
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1087 AAGTGGTGTATGACCAACGCGCTCTCGACATTCTGCATGCGGTCATGTCTCTATCTG 1146
Qy 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1147 GCGAATGCGCGCAAGCTGGCGACCGTCTGATCGTGGCGGTCAACAGCGATGCTCCACC 1206
Qy 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1207 AAACGCTGTGAAGGCGAAGCCCGCGGTGAACCGCTGGAAACCGGATGATGCTGCTG 1266
Qy 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1267 GGGCGCTGGAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1326
Qy 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTyrLysProGluIle 440
Db 1327 ATCGCTGTCATCTCCCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1386
Qy 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuValLeuAsp 460
Db 1387 GCGGCGCGAAGAGCTGTGGCGAAGCGGTGGCGAAGTGTGCTGCTGCTGCTGCTGCTG 1446
Qy 461 GlyCysSerThrAsnIleLysLysIleGlnGlnAsp 474
Db 1447 GGCTGCTCAACTACCAATATTATTAAAGAAATACAAAAAGAT 1488
```

RESULT 13

```
US-09-543-681A-2462
; Sequence 2462, Application US/09543681A
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2462
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2462
```

```
Alignment Scores:
Pred. No.: 5,51e-170 Length: 1476
Score: 2007.00 Matches: 389
Percent Similarity: 91.33% Conservative: 43
Best Local Similarity: 82.24% Mismatches: 41
Query Match: 83.87% Indels: 0
DB: 21 Gaps: 0
```

```
US-09-912-020-325 (1-477) x US-09-543-681A-2462 (1-1476):
Qy 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
Db 52 ATGAAAGTAAAGCTCCCGGATTTTAAATAAAGCAATATGCTGGT73TCGGTGACGTCATG 111
Qy 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 112 TTAGACCGTTATTGGTATGGTCTACAAAGTCGGATCTCACCAGAACACCTGTGCTGTG 171
Qy 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
Db 172 GTAAAGTAGATACCATAGAGAGCGTCCGGGTGGTGCAGCTAACCTTGCATGAATATT 231
Qy 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 232 GCTCTTTAGGGGTAAATTCGTTTGGTGGGTAAACCGGCATCGATCAAGCGCAAAA 291
Qy 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 292 GCATTAAGCAATACATTAAAGTCAGGTTAATGTCGGTGTGATTCTCTCTCTATTCCGACA 351
Qy 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 352 CATCCAACATTACTTAAATTAACGTCTCTATCAGCAACCAACACTTATCGCCTTGAT 411
Qy 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 412 TTTGAAGAGGATTTAGCAATGCTCTCCAGAGCCAAATTTATGAGCGTATTCAACAAGCA 471
Qy 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 472 TTACCTCTCAATGGTGCCTTAATTTATCGGATTATGCTAAAGGTGCATTTCTCACGTA 531
Qy 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 532 CAGCAAAATGATCCAAATAGCTAAAGAACAGGTGTCGCGGTACTAATCGATCTCTAAAGC 591
Qy 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 592 AGTGATTTTGAACGCTACCGTGGTCCACACATTTAAACCCGATATGTCAGAGTTTGA 651
Qy 201 AlaValValGlyLysCysLysThrGluGluIleValGluArgGlyMetLysLeuIle 220
Db 652 CAAAGTGGTGGCGTTTGTGCATACACATGATGAGCTAGTAGAAAAAGCGCACTCAACTAGTA 711
Qy 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 712 AAAGATTTAGCACTTGTATGCACATATTGATCTCTCTTCAGAGCGTSGCATGATTATTA 771
Qy 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 772 CAAGCCATGAAGCCCATTCATTTACCACTCAAGCACAAGAAGTTTATGATGTGACA 831
Qy 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 832 GGGCGCGGAGATACGTTATCGCGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 891
Qy 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValValValValValVal 300
Db 892 GCGGAAGCGTGGCTTCTTGCACACGCTCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 951
Qy 301 SerThrValSerProIleGluLeuGluAsnAlaValArgAlaGlyArgAlaAspThrGlyPhe 320
Db 952 TCAACGGTTTCCTCTATTGAGCTTGAATGCTATTTCGCGCGCGTGTGATTAACGGCTTT 1011
Qy 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaAlaAlaArgLysArgGlyGlu 340
Db 1012 GGTATTATGGAAGAGTCAACTCAACAAACGCGGTGCTTTAGCGCGTCAACGCTGGTGAG 1071
Qy 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 1071 GGTATTATGGAAGAGTCAACTCAACAAACGCGGTGCTTTAGCGCGTCAACGCTGGTGAG 1071
```

```
Db 1072 CGCATAGTATGACAAATGGTGTGTTTGATATCTTCATGCTGGCGATGTTAGTTATTTA 1131
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 1132 GCTAATGCAGTAAATGGTGGTATCGCTGATTTGTCGGTAAATAGTATGATCAACT 1191
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 1192 AAACGATTAAAGGTGAAGTCGCCAGTCAACCCATTAGACGCAACGAATGACCGTATTA 1251
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 1252 GGGCGGTATGGGCGAGTGTATGGGTGTCGGCTTTGAAGAAGATACCCACAGCGTTTA 1311
QY 421 IleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTyrLysProGluGluIle 440
Db 1312 ATTGCATCAGTATGCTGTATATTTAGTAAAGGTGGCGATTTATAACCCGAGACATT 1371
QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
Db 1372 GCTGGTAGTAAGAGGTGTGGGCTGCGGTGGTGAAGTGAAGTACTTAATTTTGAAGAT 1431
QY 461 GlyCysSerThrThrAsnIleIleLysLysIleGlnGln 473
Db 1432 GGTATTTCCAACTAATATTTATTAATGCCATTTAAAG 1470

RESULT 14
US-09-897-516-1461/C
; Sequence 1461, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Todorov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 1461
; LENGTH: 27693
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (877)..(1428)
US-09-897-516-1461

Alignment Scores:
Pred. No.: 5,42e-165 Length: 27693
Score: 1968.00 Matches: 383
Percent Similarity: 90.25% Conservative: 43
Best Local Similarity: 81.14% Mismatches: 46
Query Match: 82.24% Indels: 0
DB: 33 Gaps: 0

US-09-912-020-325 (1-477) x US-09-897-516-1461 (1-27693)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValGlyAspValMet 20
Db 22544 ATGAACGTACACTCCCTGATTTTCACCGTGCAGGTGTTTGGTTGTGGCGATGATG 22485
QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProVal 40
Db 22484 TTAGATGCCTATTGGTAGCGCCCAACAGCGCATTTCCACCGAAGCGCGTTCCTGTT 22425
QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
```

```
Db 22424 GTTAAAGTTGACACCATCGAAGAGCGTCCGGGGAGCAGCTAACGTTGCTATGAATATC 22365
QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
Db 22364 GCATCATTTGGCGCTTAATTCACATTTAGTCGGGTAAACGGGTATTGATGATGAGCTCGT 22305
QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
Db 22304 GCATTAAGCGAAACCTGACGCGTGAAGTCCGGTGTGATTTCTTCCTCCGTTTCAACC 22245
QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp 120
Db 22244 CATCCGACTATTACGAAATTCATTTCTTCGCTAATCAGCAACTGATCCGCTCTGAT 22185
QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
Db 22184 TTTGAAGAAGGCTTCCAGAACCTAGATGCTCAGCCAATGCTGGAGAAATTCAGCAATCT 22125
QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
Db 22124 CTGCCAATATCGTGCACCTGGTATTGTCGATTAACGCTAAAGCGCGCTGAGTCAGTT 22065
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 22064 CAGGCAATGATTAAAGCTGGCAATGATGCTGCTGCCAGTTCTGATGACCCCTAAAGGC 22005
QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
Db 22004 CAGCACTTTGAGCGTTATCGTGGCGCAACCTGCTGATCCGAATATGTCAGAAATTTGAA 21945
QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
Db 21944 GCTGTGGTGGTTCATTGCCAAGATGATGATGAAGTGGTGGCGCAAGGTACGAGCTGTA 21885
QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
Db 21884 CAAGATCTGGAACCTGAAAGCACTGTTAATCACCGTTTCAGAGGCTGCGATGCTGCTG 21825
QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
Db 21824 CGTGTGGGCGAGCTCTTTACATCTGCCAATCAGCGCGCAGGAAGTTTGTGATGTAACG 21765
QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
Db 21764 GCGCGCGGTGACAGCTAATTTGTTCTTGGCAACGCAATTCAGCGCGTGAAGCCACTG 21705
QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
Db 21704 GCTGAAGCGCTGTTATCTGCTAATGCGGCTGCGGAGTGTAGTCGCTAACTGGGTACC 21645
QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
Db 21644 TCCACTGTTTCCCGGTGGAAATTTGGAAATGCGGTGCGTGGCGGTGCAGAAACAGCGTTT 21585
QY 321 GlyValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlu 340
Db 21584 GGTGTGATGAATGAATCACGCTGAAAGAGTGTGTTGAGCAGCGCGCTGAGCGCGGTGAA 21525
QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 21524 CGTATCGTAATGACGAACGCTGTTTCGATATTCTCCATCGGGGCGACGATATCTATTG 21465
QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 21464 TCCAAATGCCGCTAACTGGCGATCGCTAATTTGGCCGCTTACACTGATGCTCAACC 21405
QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 21404 AGACGCTGAAAGGTGAAGCGCAGCACTTAATCCGCTGGAACAGCGCAATGATTGTACTG 21345
QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
```


Db 21344 TCTGCACTGGAGTCAGTGGACTGGGTGTTGTCATTTTGAAGAAGATACACCGCAGCGCTA 21285
 QY 421 ILeAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGluGluIle 440
 Db 21284 ATTGCAGGCAATTTGGCTGATATTTGGTCAAGGGCGGGATTTAAGCCGAAGAGATC 21225
 QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 Db 21224 GCTGGCAGTCAGGAAGTCCTGGGCGCTGGTGGTGAAGTACAGGTGCTGAATTTGAAGAC 21165
 QY 461 GlyCysSerThrThrAsnIleLysLysIleGln 472
 Db 21164 GGGATCTCAACGACCAATATTTAAGGGGATAAAA 21129

RESULT 15

US-09-897-516-1462/c
 ; Sequence 1462, Application US/09897516
 ; GENERAL INFORMATION:
 ; APPLICANT: Corbin, David R.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Huesing, Joseph E.
 ; APPLICANT: Krasomil-Osterfeld, Karina C.
 ; APPLICANT: Malvar, Thomas M.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Spiridonov, Sergei
 ; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
 ; FILE REFERENCE: 38-21(51847)B
 ; CURRENT APPLICATION NUMBER: US/09/897,516
 ; CURRENT FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/215, 161
 ; PRIOR FILING DATE: 2000-06-30
 ; NUMBER OF SEQ ID NOS: 8409
 ; SEQ ID NO 1462
 ; LENGTH: 27693
 ; TYPE: DNA
 ; ORGANISM: xenorhabdus sp.
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1739)..(1870)
 US-09-897-516-1462

Alignment Scores:

Pred. No.: 5,42e-165 Length: 27693
 Score: 1968.00 Matches: 383
 Percent Similarity: 90.25% Conservative: 43
 Best Local Similarity: 81.14% Mismatches: 46
 Query Match: 82.24% Indels: 0
 DB: 33 Gaps: 0

US-09-912-020-325 (1-477) x US-09-897-516-1462 (1-27693)

QY 1 MetLysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMet 20
 Db 22544 ATGAAGCTAGCACTCCCTGATTTTACCGTGCAGGTGTTTGGTGTGGCGATGTCATG 22485
 QY 21 LeuAspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProVal 40
 Db 22484 TTAGATCGCTATTGGTACGCCCAACAGCCGCAATTCACCGGAAGCGGGTTCCTGTT 22425
 QY 41 ValLysValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIle 60
 Db 22424 GTTAAAGTTGACACCATCGNAGAGCGTCCGGCGGAGCAGCTAACGTTGCTATGAATATC 22365
 QY 61 AlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArg 80
 Db 22364 GCATCATTTGGCGCTAATTCACATTTAGTGGCTTAACGGGTATTGATGATGCAGCTCGT 22305
 QY 81 AlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValProThr 100
 Db 22304 GCATTAAAGCAAAACATGACGCGTGAAGTCCGGGTGATTTGTTCCGTTTCAACC 22245
 QY 101 HisProThrIleThrLysLeuArgValLeuSerArgAsnGlnLeuIleArgLeuAsp 120

Db 22244 CATCCCACTATTACGAAATTCATGTTCTTCGGTAATCAGCAACTGATCCGCTCGCAT 22185
 QY 121 PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAla 140
 Db 22184 TTTGAAGAAGGCTTCCAGAACGTAGATGCTCAGCAACTGCTGAGAAAAATTCAGCAATCT 22125
 QY 141 LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerVal 160
 Db 22124 CTCCCAATATCCGTCACCTGGTATTGTCGATTACGTAAGCGCGCTGAGTCAGGTT 22065
 QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
 Db 22064 CAGGCAATGATTAAGCTGGCAAAATGATGCTGGTGTCCACGTTCTGATTGACCCCTAAAGGC 22005
 QY 181 ThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGlu 200
 Db 22004 CACGACTTTGAGCGTTATCTGTCGCAACCCCTGCTGACTCCGAATATGTCAGAAATTTGAA 21945
 QY 201 AlaValValGlyLysCysLysThrGluGluGluIleValGluArgGlyMetLysLeuIle 220
 Db 21944 GCTGTGGTGGTCAATTCGAAAGATGATGATGAACTGGTGGCAAGGTACGAAAGCTGGTA 21885
 QY 221 AlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeu 240
 Db 21884 CAAGACTCTGAACTGAAAGCACCTGTTAATCACCCGTTTCAGACGCTGCGATGCTGCTG 21825
 QY 241 GlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThr 260
 Db 21824 CGTGTGGCGCAGCTCTCTTTACATCTGCCAACTCAGCGCAGGAAGTTTTTGATGTAAAC 21765
 QY 261 GlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeu 280
 Db 21764 GGCCTGGGTCACACGGTAATTGTTCTGGCAACGGCAATTCGACCGGTGCGAAGACGGCTTT 21705
 QY 281 GluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr 300
 Db 21704 GCTGAAGCCCTGTTATCTGCGTAATCGCGCTGCGGAGTGTAGTCGCTAAACTGGGTACC 21645
 QY 301 SerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPhe 320
 Db 21644 TCCACTGTTTCCCGGTGGAAATTTGGAATTCGCGTGGCGGTGCGAAGACAGGCTTT 21585
 QY 321 GlyValMetThrGluGluLeuLysLeuAlaValAlaAlaAlaAlaArgLysArgGlyGlu 340
 Db 21584 GGTGTGATGATGAATGAATCACCCCTGAAGAAGTTGTTGACGAGCGCGCTGACCGCGTGAA 21525
 QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
 Db 21524 CGTATCGTAATGACGAACGCTGTTTCGATATTCTCCATCGGGGCGACGTATCTTATTG 21465
 QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
 Db 21464 TCCAATGCCGTAACTAGGCGATCGCCTAATTTGTGCGCTTAACTAGTGTGCTCAACC 21405
 QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
 Db 21404 AGACGCTGAAGAGTGAACCGCACCGAGTAAATCCGCTGGAACGCAATGATTGTACTG 21345
 QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluAspThrProGlnArgLeu 420
 Db 21344 TCTGCACTGGAGTCAGTGGAGTGGTGTGTCATTTGAAGAAGATACACCGCAGCGCTA 21285
 QY 421 IleAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGluGluIle 440
 Db 21284 ATTGCAGGCAATTTGGCTGATATTTGGTCAAGGGCGGGATTTAAGCCGAAGAGATC 21225
 QY 441 AlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPheGluAsp 460
 Db 21224 GCTGGCAGTGAAGAAGTCTGGGCGCTGGTGGTGAAGTACAGGTGCTGAATTTGAAGAC 21165
 QY 461 GlyCysSerThrThrAsnIleLysLysIleGln 472

Db 21164 GGGATCTCAAGGACCAATATTATTAAAGCGATAAAA 21129

Search completed: November 27, 2002, 01:55:53
Job time : 3642 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 26, 2002, 21:16:58 ; Search time 112 Seconds
(without alignments)
1876.398 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERAGVMVGVDM.....FEDGCTTIIKIQDQKKG 477

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 288240 seqs, 220289906 residues

Total number of hits satisfying chosen parameters: 576480

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09912020/runat_25112002_091430_24333/app_query.fasta_1.647
-DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=p2n.rnpn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09912020.ecgn_1_1_54 -runat_25112002_091430_24333
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending_Patents_NA_New.*

1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	190.5	8.0	US-09-134-000C-1119	Sequence 1119, Ap
2	170.5	7.1	US-10-264-213-15	Sequence 15, Appl
3	165.5	6.9	US-10-092-411A-1684	Sequence 1684, Ap
4	150	6.3	US-10-092-411A-2723	Sequence 2723, Ap
5	145	6.1	PCT-US02-32727-56	Sequence 56, Appl
6	143.5	6.0	PCT-US02-32727-107	Sequence 107, App
7	140	5.9	PCT-US02-32727-176	Sequence 176, App
8	137.5	5.7	US-10-092-411A-2746	Sequence 2746, Ap
9	133	5.6	PCT-US02-32727-132	Sequence 132, App
10	131	5.5	US-09-134-000C-1139	Sequence 1139, Ap
11	131	5.5	US-10-264-237-1121	Sequence 1121, Ap

12	123	5.1	61557	1	PCT-US02-32727-16	Sequence 16, Appl
13	120.5	5.0	810	5	US-09-134-000C-1809	Sequence 1809, Ap
14	120	5.0	12023	6	US-10-264-213-100	Sequence 100, Appl
15	119.5	5.0	1877	5	US-09-724-676-45787	Sequence 45787, A
16	118.5	5.0	1811	5	US-09-724-676-45788	Sequence 45788, A
17	118.5	5.0	1875	5	US-09-724-676-45791	Sequence 45791, A
18	116.5	4.9	8784	1	PCT-US02-32727-27	Sequence 27, Appl
19	116	4.8	1748	5	US-09-724-676-45794	Sequence 45794, A
20	115.5	4.8	6360	6	US-10-212-962-2	Sequence 2, Appl
21	115	4.8	1682	5	US-09-724-676-45795	Sequence 45795, A
22	115	4.8	1811	5	US-09-724-676-45796	Sequence 45796, A
23	115	4.8	1811	5	US-09-724-676-45789	Sequence 45789, A
24	115	4.8	1875	5	US-09-724-676-45792	Sequence 45792, A
25	114	4.8	1616	5	US-09-724-676-45797	Sequence 45797, A
26	114	4.8	1745	5	US-09-724-676-45790	Sequence 45790, A
27	114	4.8	1809	5	US-09-724-676-45793	Sequence 45793, A
28	113	4.7	1825	5	US-09-724-676-21637	Sequence 21637, A
29	112.5	4.7	29112	1	PCT-US02-32727-104	Sequence 104, Appl
30	111.5	4.7	29379	1	PCT-US02-32727-5	Sequence 5, Appl
31	110	4.6	966	6	US-10-092-411A-1063	Sequence 1063, Ap
32	110	4.6	11220	6	US-10-271-889-32	Sequence 32, Appl
33	110	4.6	36778	6	US-10-271-889-48	Sequence 48, Appl
34	110	4.6	38506	6	US-10-201-363-1	Sequence 1, Appl
35	110	4.6	38506	6	US-10-160-539A-19	Sequence 19, Appl
36	109.5	4.6	35829	1	PCT-US02-32727-68	Sequence 68, Appl
37	109.5	4.6	56050	1	PCT-US02-32727-44	Sequence 44, Appl
38	108	4.5	972	5	US-09-134-000C-649	Sequence 649, Appl
39	108	4.5	29559	1	PCT-US02-32727-41	Sequence 41, Appl
40	108	4.5	37286	1	PCT-US02-32727-17	Sequence 17, Appl
41	107	4.5	3225	5	US-09-134-000C-2685	Sequence 2685, Ap
42	106	4.4	957	5	US-09-134-000C-2976	Sequence 2976, Ap
43	106	4.4	6107	5	US-09-751-708A-102	Sequence 102, App
44	105.5	4.4	981	5	US-09-134-000C-2202	Sequence 2202, Ap
45	105	4.4	1907	5	US-09-724-676-21636	Sequence 21636, A

ALIGNMENTS

RESULT 1

US-09-134-000C-1119
; Sequence 1119, Application US/09134000C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1119
; LENGTH: 1590
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1119

Alignment Scores:
Pred. No.: 1.74e-08 Length: 1590
Score: 190.50 Matches: 31
Percent Similarity: 40.48% Conservative: 53
Best Local Similarity: 24.47% Mismatches: 126
Query Match: 7.96% Indels: 71
DB: 5 Gaps: 11

US-09-912-020-325 (1-477) x US-09-134-000C-1119 (1-1590)

Qy	13	ValMetValValGlyAspValMetLeuAspArgTyrTrpTyr-----	26
Db	655	ATGATTGTAGTGGTGGTAAACGTTGATATCTGTATTCACAAATTTACCACAT	714
Qy	27	--GlyProThrSerArgIleSerProGluAlaProValProValValAsnThr	45

```
Db 715 AACGGCAAAACAATTTTGTCTAGCTATGCTAAATTCGCC----- 756
QY 46 IleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAla 65
Db 757 -----GGCGAAAGGATTAAATCAAGCTGTTCGCTTAACCAACATGGGCCAT 804
QY 66 AsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSer 85
Db 805 CAGGCCACCTTAATTTGGTTTGGCTCAGACACAGATCGGAATTACTTATATATAAGAA 864
QY 86 LeuAlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThr 105
Db 865 TTAGAAAAATACACCGTTACAGAGAT-----GGTATCACCA 900
QY 106 LysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp----- 120
Db 901 CGTATTCAGATGACTGAGCTGGTCAAGCTATATTATGAGACACAGCGGGAGCTCC 960
QY 121 -----PheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArg 136
Db 961 ATGATTTCATCTTACCTGTGTGCAACACACAGCGCTTACGCTTAAATAAATCGCACAGCAA 1020
QY 137 IleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAla 156
Db 1021 AAACACCTATTATGGATGCCAGTTTGTCTCATTTCAAAACAGAAAT-----CCT 1071
QY 157 LeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIle 176
Db 1072 TTGCTCTCTGAGAGAAAGCTTGTGAATCGCCACATTCAGGGTACGATATTATTTA 1131
QY 177 AspPro-----LysGlyThrAspPheGluArg 185
Db 1132 AAACAGCTGCTATTCATCATATATCCAGTGAATATACAGAAAAAGGTGATTTT----- 1185
QY 186 TyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLys 205
Db 1186 -----TTCATTCCCAATGAAGATGAATATTGGAACTTCAACCCAGAT 1227
QY 206 CysLysThrGluGluGlu-----IleValGluArgGlyMetLysLeuIleAla 221
Db 1228 ACTGGTACATTAGAGAGAAAGCGCGTACTTCTTAGAATGGCGTAAATAATGTC--- 1284
QY 222 AspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMetSerLeuLeuGln 241
Db 1285 -----ATTGTTACTTTAGGAAAAAAGCGGTG---TTATTAAAA 1320
QY 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGly 261
Db 1321 ACACCCCAAGTGTGCCACTATTTCCCTGCAACAGAAAAATATAGCTGTGGATAGCAGCTGGT 1380
QY 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 1381 GCCAGTACAGCTTTATCAGCGCGCTCGCTCTATCTTCGAAAGGTTATCCGACTGAA 1440
QY 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThr--- 300
Db 1441 GCAGCCATTCAATAGCAATTCAGGCTGCAGGATTTTCAGTTTCTTAAAGAAGGGTGATT 1500
QY 301 ---SerThrValSerProIleGluLeuGluAsn 310
Db 1501 GATTCTTTAGTTGATCAGCTACGTTTGGAAAT 1533
```

RESULT 2

US-10-264-213-15/c

; Sequence 15, Application US/10264213

; GENERAL INFORMATION:

; APPLICANT: Glenn, Matthew

; APPLICANT: Havukkala, Ilkka J

; APPLICANT: Lubbers, Mark William

; APPLICANT: Dekker, James

; TITLE OF INVENTION: Polynucleotides, materials incorporating

; TITLE OF INVENTION: them, and methods for using them.

```
; FILE REFERENCE: l1000.1043c3
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 7210
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-15
```

```
Alignment Scores:
Pred. No.: 8.37e-06 Length: 7210
Score: 170.50 Matches: 74
Percent Similarity: 41.99% Conservative: 57
Best Local Similarity: 23.72% Mismatches: 144
Query Match: 7.12% Indels: 37
DB: 6 Gaps: 14
```

US-09-912-020-325 (1-477) x US-10-264-213-15 (1-7210)

```
QY 13 ValMetValValGlyAspValMetLeuAspArgTyrTrpTyrGlyProThrSerArgile 32
Db 5388 GTTGTCTGTATCGCAGTATCAATGTGGATGCTATTTTACAC-----ATTCACGACTG 5335
QY 33 SerProGluAlaProValProValValLysValAsnThrIleGluGluArgProGlyGly 52
Db 5334 ---CCACAG---CCGGGAGAAACAATCAATGATGCTTTTCTAAAGCTCGCGTGGT 5281
QY 53 ---AlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeuValGly 71
Db 5280 AAGGGGCCAATCAAGCGTTGCGGCGCGATCTGCGCTAGACCAATTCATCGGG 5221
QY 72 LeuThrGlyIleAspAlaAlaArgAla-----LeuSerLysSerLeu 86
Db 5220 CGCGTCCGG---GATGATGCCAATGCTGCTTCAGGCGTGGCGAGTGTGTAATAATCAA 5164
QY 87 AlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLys 106
Db 5163 ATTGATACCAATATGTTGGACACCCAGACACGGAACCCGCCAGGCGTATATTCTG 5104
QY 107 LeuArgValLeuSerArgAsnGlnGlnLeuIleArg-----LeuAspPheGluGly 124
Db 5103 CTCACAGATCGGGCAAAACTCGATCATTCATCAACATGGCGCAACTTTGAA----- 5050
QY 125 PheGluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerile 144
Db 5049 -----CTGACGCCAGCGGATGTCAGCGG-----GCCACT 5020
QY 145 GlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeu-----AlaSerVal 160
Db 5019 GGCTTGATTCAATCGCGTGACTTGTGTGCTGAACCTGGAAACGCGGTAGCGCTACA 4960
QY 161 GlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGly 180
Db 4959 GCCGAAGCATTCAAAATTCGCGGTGAAGTAACCACTCTTAAACCTGCACCG 4900
QY 181 ThrAspPheGlu-----ArgTyrArgGlyAlaThrLeuLeuThrProAsnLeu 196
Db 4899 CGCCAAAAGAACTGCCAGAGCGCTGTTGAAAATTTGATCTATATCGGCCCAACGAG 4840
QY 197 SerGluPheGluAlaValValGlyLysCysLysThrGluGluIleValGluArgGly 216
Db 4839 ACTGAAAGCGAATTGATTACCGCATTCGCGTACCGATGAAGCCAGCATGCCAGTCC 4780
QY 217 MetLysLeuIleAlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGly 236
Db 4779 GCGCGGTATTATCATCAACTCGGCATCCGGGGTGTGTCATCAGCTTGGTTGCAAGGG 4720
QY 237 MetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluVal 256
Db 4719 TCTTTTCATTTCGTGGATGGTCAGGCAACCCCTC---GTCCCGGCTTCAAGGTCAAAGCT 4663
```

Qy 257 TyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAla--- 275
Db 4662 GTGATACTACGCAGCAGGGGATACCTTTATCGCGCTTTAGTCAGCAACCTTCAGCCC 4603
Qy 276 ---AlaGlyAsnSerLeuGluGluAlaCysPheAlaAsnAlaAlaGlyValVal 294
Db 4602 GATCTCGCAAT---ATTGTCGATCGCAACTTACGCCAGCATGCCAGCTCGTTTACT 4546
Qy 295 ValGlyLysLeuGlyThrSerThrValSerProIle 306
Db 4545 GTCCAGAACCTCGCGCATTCCTCGATTCGGATG 4510

RESULT 3

US-10-092-411A-1684
; Sequence 1684, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1684
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-1684

Alignment Scores:

Pred. No.: 1,53e-06 Length: 948
Score: 165.50 Matches: 77
Percent Similarity: 39.12% Conservative: 56
Best Local Similarity: 22.65% Mismatches: 142
Query Match: 6.92% Indels: 65
DB: 6 Gaps: 13

US-09-912-020-325 (1-477) x US-10-092-411A-1684 (1-948)

Qy 2 LysValThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMetLeu 21
Db 7 AAAGTAGGTGAAGTTGAAATGAATGAAATAAAGTATGTTGTAATGGATCAACAATGTA 66
Qy 22 AspArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValVal 41
Db 67 GATAAATTTCTT-----AATGTTAAAGGTTTCCAAA---CCCGGTGAGACATTA 114
Qy 42 LysValAsnThrIleGluGluArgProGlyGly-----AlaAlaAsnValAlaMetAsn 59
Db 115 CATATTAACTACAGCTCAAAAGGAGTTGTGGGGCAAGGAGGCAATCAAGCCATAGCA 174
Qy 60 IleAlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAla 79
Db 175 GCTAGTAGATTAGCAGAGATACACATTTATCAGTAAGTTGGTAAAGATGGCAATGCT 234
Qy 80 ArgAlaLeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerValPro 99
Db 235 AACTTTATATTGGAA-----GATTTCAAAAAGCAGGT 267
Qy 100 ThrHisProThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeu 119
Db 268 ATTCAT-----ACACAATATATTTTAACTTCA 294
Qy 120 AspPheGluGlu-----GlyPheGluGlyValAspProGlnPro----- 132
Db 295 GAAAGTGAAGAAACTGGGCAAGCATTTATCAGTCTGTTGATGAAGCAGGACAAATACGATT 354

Qy 133 ---LeuHisGluArgIleAsnGlnAlaLeuSer-----SerIleGly 145
Db 355 CTGTGTTACGGTGGTGGCAATATGACATTAAGTCAACTGATGTTTCAGATGAGTGGAT 414
Qy 146 AlaLeuValLeuSerAspTyrAlaLysGlyAlaLeu-----AlaSerValGln 161
Db 415 GCGTTATTGTTGGTGCAGACTTTGTTAGCGCAGCTTGAAGTTCCATTTTTCAGCGATAGAA 474
Qy 162 GlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspPro----- 178
Db 475 CAAGCATTTAAATTTGCTGCTAAACAAAATATCACTACTGTATTAAATCTGCACCGGCA 534
Qy 179 -----LysGlyThrAspPheGluArgTyrArgGlyAlaThr 190
Db 535 ATTGAATTCCTAAGTCACTTTTAGAGTTAACTGAT----- 570
Qy 191 LeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLysCysLysThrGluGlu 210
Db 571 ATAATTATTCCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 630
Qy 211 GluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuVal 230
Db 631 AGTGATATCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 690
Qy 231 ThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetPro 250
Db 691 ACTTTAGGGGAGCAGGACGCTATTGTGCATATCAAGCAACATACAAAATG---ATTCT 747
Qy 251 ThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeu 270
Db 748 GCGTGTAAATGTAAGCAATAGATACGACAGCAGCAGGAGAGATACATTTATAGGTGCTTT 807
Qy 271 AlaAlaThrLeuAlaAlaGly---AsnSerLeuGluGluAlaCysPhePheAlaAsnAla 289
Db 808 TTAAGTGAGTTAAATAAGATTTGAGCAATTTAGAACGCTATTTCGACTTGCAAATCAA 867
Qy 290 AlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerProIleGluLeuGlu 309
Db 868 GCGTGTCTCTTAACGCTACACAGCAAGGAGCAGCAGCAAGCTTCTATACCAACACGTAAGAA 927

RESULT 4

US-10-092-411A-2723
; Sequence 2723, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 2723
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-2723

Alignment Scores:
Pred. No.: 1.23e-05 Length: 414
Score: 150.00 Matches: 47
Percent Similarity: 54.68% Conservative: 29
Best Local Similarity: 33.81% Mismatches: 49
Query Match: 6.27% Indels: 14
DB: 6 Gaps: 6

US-09-912-020-325 (1-477) x US-10-092-411A-2723 (1-414)

Db	2512	AGGTCGGCTGGAGAGATTTCCGGCCCGACGGGAGACCCCTGGCCGGAGTGGAGCGGACA	2571
Qy	48	GluArgProGlyGly---AlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn	66
Db	2572	CTGACCCCGCGGTAAAGGAGCTAACAGGCTCTTCGGCGGGCGCATGTGGCTTCAG	2631
Qy	67	AlaArgLeuValGlyLeuThrGlyIleLeuAspAlaAlaArgAlaLeuSerLysSerLeu	86
Db	2632	GTGCGATGTCTCGAGCTGTGGGAAGGACCGCGGCTGAACG---GCCCTATCCCTA	2688
Qy	87	AlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThrLys	106
Db	2689	CTATCGGAAGCGTTGGCTTGAT-----GCGGTGGCTCGCTGACGATTACGGGGG	2742
Qy	107	LeuArgValLeu-----SerArgAsnGlnGlnLeuIleArgLeuAspPheGlu	122
Db	2743	ACGCCGCTGTCATGTGTCGGGATTCGGAGAGAACTCAATTATGTCATCGCGCGCGCT	2802
Qy	123	GluGlyPheGluGlyValAspProGlnProLeuHis-----GluArgIleAsnGln	139
Db	2803	AATGGG-----ACCGTTGATGACAGACCGTCCGGGCTCAGCGCGAGSCGATCCAGAC	2856
Qy	140	Ala-----LeuSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeu	157
Db	2857	GCTGACGTGTGTGTCAAGGTGAGATTCGCGCTGACGGGATTCGCGAGTGCCTGCG	2916
Qy	158	AlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGluValProValLeu---Ile	176

Db 2917 GCGCCAAACGATTCGTTCATCTGGCA-----CCGGTGTGACGTT 2961

QY 177 AspProLysGlyThrAspPheGluArgTyrArgGlyAlaThrLeuLeuLeuThrProAsnLeu 196

Db 2962 GATCCA-----GAGGTCATTCGCATGSCCGACCGCTGGTAGTTAAC--- 3003

QY 197 SerGluPheGluAla-----ValValcIyLysCysLysThrGlu---Glu 210
||| ||||| :||| :|||

Db 3004 ---GAACATGAGCGCGCCTTGGTGGCTCTTGGCTCTTGGGATCGACGACTTCTCGTCTTGAG 3060

QY 211 GluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAlaLeuLeuVal 230

Db 3061 GATGATCCTGACGCAGCCTTGGCGGAGCTGTGAACCTTGGTSCCGCAGTGTGTGCATC 3120

Qy 231 ThrArgSerGluClnGlyMetSerLeuLeuClnProGlyLysAlaProLeuHisMetPro 250

Db 3121 ACCTGGGATCGCAGGATGCATCTCGGGGGCCGGAAGGTT CGACGCCGTGGTGGC 3180

Qy 251 ThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeu 270

Db 3181 GCGACCGTGAAGGCTGTC---GATACCGTCGGTGGGGAGATGCTTTTCGTGGTGGCGTTG 3237

Qy 271 AlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAla 290

Db 3238 GCTGCAGAACTGGCTCGTGGCCGFACTCTTGTCTGAGGTTGCCGTTTGTACCGCGGTG 3297

Qy 291 AlaGlyValValGlyLysLeuGlyThrSerThrValSerProIle-----GluLeu 308

D_b 3298 GCGACTTTGACGGTGACCAAGCCGGGGGCAAGCCTCTCTACCCAGTGCTGACGAGGTG 3357

Qy 309 GluAsnAlaValArgGly-----ArgAlaAspThrGlyPheGlyValMet 323
|||::: ||||| |||||

Db 3358 GAGAGCATTCTCGAGGAGAACACGCATGAAGAGCGGAC----- 3399

Qy 324 ThrGluGluGluLysLeuAlaValAlaAlaArgLysArgGlyGluLysValVal 343
:::

Db 3400 -----TGCTTA 3405

Qy 344 MetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAla 363

Db 3406 ATCCCAATTGTGTGCG-GCGGTGGCCAGGTTGGGCACACCCAG----- 3449

Qy 364 ArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThrLysArgLeu 383

Db 3450 -----ACCTTCGTCGTTGCTGATGCTGGGTG----- 3476

```
Qy 384 LysGlyAspSerArgProValAsn-----ProLeuGluGlnArgMetIleValLeu 400
Db 3477 -----CCTATTCCTCATGAGTCCCGGTGATGACCTCGCTGCTGCTG 3521
Qy 401 GlyAlaLeuGluAlaValAspTrpValSerPheGluGluAspThrPro-----Gln 418
Db 3522 GGA-----ACTCCGAGGTTCAG 3539
Qy 419 ArgIleuIleAlaGlyIleLeuProAspLeuValLysGlyAspTyrLysProGlu 438
Db 3540 GAGGTTCGAGCGCATCTCCGACGAGTGTGTCGGACGAGCACCATTGCTCACCAG 3599
Qy 439 GluIleAlaGlySerLysGluValTrpAlaAsnGly-----GlyGluValLeuValLeu 456
Db 3600 GCTCTCGGGCATGACCGGAATCTTGGGTGAGAGAACGGATTGAAGAGGTTTCACACGGTC 3659
Qy 457 AsnPheGluAsp 460
Db 3660 AGTCATGAGGAC 3671

RESULT 6
PCT-US02-32727-107/c
; Sequence 107, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Wang, Siging
; APPLICANT: Jen, Shylan
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darriack
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 107
; LENGTH: 21263
; TYPE: DNA
; ORGANISM: Propionibacterium
PCT-US02-32727-107

Alignment Scores:
Pred. No.: 0.00956 Length: 21263
Score: 143.50 Matches: 75
Percent Similarity: 37.75% Conservative: 39
Best Local Similarity: 24.83% Mismatches: 123
Query Match: 6.00% Indels: 65
DB: 1 Gaps: 12

US-09-912-020-325 (1-477) x PCT-US02-32727-107 (1-21263)

Qy 13 ValMetValValGlyAspValMetLeuAspArgTyrTrpTyr-----GlyProThr 29
Db 13212 ATCGCGGTGTGTGGTCCCAATATGTGGACCTCATCTCGTACATTCATCGCATGCCATCA 13153
Qy 30 SerArgIleSerProGluAlaProValProValValLysValAsnThrIleGluGluArg 49
Db 13152 GACGGTGAACCGTCGAGGCTCC-----GAATTCGGA 13120
Qy 50 -----ProGlyGlyAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsn 66
Db 13119 ATGGGATCGCGCAAGGTGCGAAGTAGTCAGTGTGCTCGGTGCAGAG 13060
Qy 67 AlaArgLeuValGlyLeuThrGlyIleAsp----- 76
```

```
Db 13059 GTCGTGATGGTCACCCGAGTGGCAACGATGTGTTCCGCGACACCACTTTGACAACTTT 13000
Qy 77 -----AspAlaAlaArgAlaLeuSerLysSerLeuAlaAspValAsnVal 91
Db 12999 CGAAAAACGGTATCGATACCTACCATGTGTG---CGTACCGATGCGAGCTCCGGGTG 12943
Qy 92 LysCysAspPheValSerValProThrHisProThrIleThrLysLysArgValLeuSer 111
Db 12942 GCACCGATCTTTCGTCGATCCAGAAATCGCATCACTCATC-----CTCATCAAGGGA 12889
Qy 112 ArgAsnGlnGlnIleAlaArgLeuAspPheGluGluGlyPheGluGlyValAspProGln 131
Db 12888 GCTAACCGTCTAGCTGAGTCTCGTGTGATCTCAGGGTGC----- 12850
Qy 132 ProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeuValLeuSerAsp 151
Db 12849 -----CGACAAGACATTCGACAGTCCAACTCATCATCTGTCGAG--- 12811
Qy 152 TyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGly 171
Db 12810 ---CTGGAGATTCCGTTGGAGACGGTCTACGCCAGATTGATTAGGGGAGCTACTGGGG 12754
Qy 172 ValProValLeuIleAspProLysGlyThr-----AspPheGluArgTyrArg 187
Db 12753 ATCCCAATACTCTCAATCCGCTCCGCTCGCACCAGAACTGGATCTTGAGCGAATCCGC 12694
Qy 188 GlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValIleGlyLysCysLys 207
Db 12693 GGAATTGAGTCTTTCATCGCGAAGAGTCCGAGCTTGAACATCAACGGGGATGCCAGTA 12634
Qy 208 ThrGluGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGluLeuSerAla 227
Db 12633 GACACCTCGATGATATCGGCAAGCTACTGATGTGCTGGGAGCGGGATCAGGAAT 12574
Qy 228 LeuLeuValThrArgSerGluGlnGly---MetSerLeuLeuGlnProGlyLysAlaPro 246
Db 12573 ATCATCGTAACCTTTGGGCTCGCTGCGATGTGGGCCACGACAGAGGACGCAAAATC 12514
Qy 247 LeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGlyAlaGlyAspThrVal 266
Db 12513 ATCAAGGCCCGC-----GTTGTTCAAGCCGCTGGACACGACAGGTGCTGGCGATCCCTC 12460
Qy 267 IleGlyValLeuAla-----AlaThrLeuAlaAla 276
Db 12459 ATAGGCTGCTTTGCTAAAGATGGTAGATCACGGGACGTGATAGCTCCGATTCTGCA 12400
Qy 277 GlyAsn 278
Db 12399 GGAAC 12394

RESULT 7
PCT-US02-32727-176
; Sequence 176, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siging
; APPLICANT: Jen, Shylan
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darriack
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
```

; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 176
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Propioni acnes
PCT-US02-32727-176

Alignment Scores:

Pred. No.: 0.00052 Length: 1434
Score: 140.00 Matches: 102
Percent Similarity: 35.23% Conservative: 59
Best Local Similarity: 22.38% Mismatches: 166
Query Match: 5.85% Indels: 133
DB: 1 Gaps: 20

US-09-912-020-325 (1-477) x PCT-US02-32727-176 (1-1434)

```
QY 35 GluAlaPro-----ValProValValLysValAsnThrIleGluGluArgPro 50
D 328 GAAGCACCTGCCCCACACCGCCTGCGCGTTCGCGTGATACGGCGCGCGGACCG 387
QY 51 GlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAlaAspVal 70
D 388 TCGTCAGGTTGTCGAGGCTGACATCGACACTGTGCACCAAGCAGGATCAGAAT-ATC 446
QY 71 GlyLeuThrGlyIleAspAsp---AlaAlaArgAlaLeuSerLysSerLeuAlaAspVal 89
D 447 GGTGCGCCCGACGTGAGGATGGCCCTTTCTCGCCCTT-----CTCAGGAGTTG 497
QY 90 AsnValLysCysAspPheValSerValProThrHisProThrIleThrLysLeuArgVal 109
D 498 GGAATCCGGTGTGAT-----CCTCAGCTCTCGCATCTCGCGTGTG 536
QY 110 LeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPheGluGlyValAsp 129
D 537 AGCACCGCTCGATGCG-----TTCGAAACCGT-----GGTATTCOC 575
QY 130 ProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIleGlyAlaLeu----- 147
D 576 ACCAAGCAGAGGCTGGAGTTTGTAGTGACGCCATCTTGAAGTTGTGTGTCACCGACTAT 635
QY 148 -----ValLeuSerAspThrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIle 164
D 636 TTGTACCGGCGCTTTCGCGATAGCTGAAGCCAGCTAGCCAAAGTTACGGTGGTGT 695
QY 165 -----GlnLeuAlaArgLysAlaGlyValProValLeuIleAsp 177
D 696 GTGAGTACCGTCTCGTAGCAGACGTGGCAGCTCACTCGGATGTG--GTCCTC----- 747
QY 178 ProLysGlyThrAspPheGluArgTyArgGlyAlaThrLeuLeuThrProAsnLeuSer 197
D 748 -----GGATCAAAATGGCAAGGAGAGGAATTCGTGCTT----- 783
QY 198 GluPheGluAlaValGlyLysCysLysThrGluGluGluIleValGluArgGlyMet 217
D 784 -----GAGGATGATCTCGACGAGCTTTG 807
QY 218 LysLeuIleAlaAspTyrgluLeuSerAlaLeuLeuValThrArgSerGluGlnGlyMet 237
D 808 CGCGAGCTGTTGAACCTTGGTGCCCGCAGTGTGTGTCATCACCCTTGGGATCGCAGGATGC 867
QY 238 SerLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyr 257
D 868 ATGTCGCGCGCGCGGAGGTTTCGAGCGCGTGGTGGCGGACCGCTGAAGGCTGTC--- 924
QY 258 AspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaGly 277
D 925 GTACCGCTCGGTGCGGAGATGCTTTCTGTGTGCTGGCTGCGAAGTCTCGTGTGC 984
QY 278 AsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyValValGlyLys 297
D 985 CGTACTCTTGTGCGAGGTTGCGGTTTTTGTACCGCGCTGGCAGACTTTTGACGGTGACCAAG 1044
```

```
QY 298 LeuGlyThrSerThrValSerProIle-----GluLeuGluAsnAlaValArgGly--- 314
D 1045 CCGGGGCGCAAGCCTCTACCCAGTGTGTGACGAGGTGGAGAGCATTCGTCGAGAGAA 1104
QY 315 -----ArgAlaAspThrGlyPheGlyValMetThrGluGluLeuLysLeu 330
D 1105 CACGCATGAAGAGAGCGGAC----- 1125
QY 331 AlaValAlaAlaAlaArgLysArgGlyGluLysValValMetThrAsnGlyValPheAsp 350
D 1126 -----TGTTPAATCCCAATTTGTGTGCG-GCG 1151
QY 351 IleLeuHisAlaGlyHisValSerTyrrLeuAlaAsnAlaArgLysLeuGlyAspArgLeu 370
D 1152 GTGGCAGGTTGGGCGACACCCAG----- 1175
QY 371 IleValAlaValAsnSerAspAlaSerThrLysArgLeuLysGlyAspSerArgProVal 390
D 1176 ---ACCTTCGTGCTGTGCTGATGCTGGTGTG-----CCGATT 1208
QY 391 Asn-----ProLeuGluGlnArgMetIleValIleValIleGlyAlaLeuAlaValAsp 407
D 1209 CCTCATGAGGTCCTGGTGTGATTCACCTCGCTGTGTGCTGGGA----- 1250
QY 408 TrpValValSerPheGluGluAspThrPro-----GlnArgLeuIleAlaGlyIleLeu 425
D 1251 -----ACTCGAGGTTCCAGGAGGTTTTCACGCCATCCTC 1286
QY 426 ProAspLeuLeuValLysGlyGlyAspTyrrLysProGluGluIleAlaGlySerLysGlu 445
D 1287 GACGAAGTGTGTGCGACGAGCCACGATGTCTCAGAGGCTCTCGGGCATGAACCGAA 1346
QY 446 ValTrpAlaAsnGly-----GlyGluValLeuValLeuAsnPheGluAsp 460
D 1347 TCTTGGGTGAGAGACGGATTGAAGAGTTTCACAGGTCAGTCATGAGGAC 1397
```

RESULT 8

US-10-092-411A-2746
; Sequence 2746, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 2746
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-2746

Alignment Scores:
Pred. No.: 0.000493 Length: 939
Score: 137.50 Matches: 58
Percent Similarity: 43.28% Conservative: 58
Best Local Similarity: 21.64% Mismatches: 111
Query Match: 5.75% Indels: 41
DB: 6 Gaps: 12

US-09-912-020-325 (1-477) x US-10-092-411A-2746 (1-939)

```
QY 51 GlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAlaAsnAlaArgLeuVal 70
D 127 GTAAGGGAATCAATGATCAAGAGTACTATAACATTTAGATTGTGATTCTACAGCCTTA 186
```

```
Qy 71 GlyLeuThrGlyIleAspAlaalaArgAlaLeuSerLysSerLeuAlaAspValAsn 90
Db 187 GGATTTTCAGGA--GGTTTCCCTGGAGATTATTGCTCAACACACTTGAGGACAGTAAT 243
Qy 91 ValLysCysAspPheValSerValProThrHisProThrIleThrLysLeuArgValLeu 110
Db 244 ATCCAATCAGATTTTGTTCAGTAGATGAGGAT-----ACGCGTATTAAATGTAATA 294
Qy 111 SerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPheGluGlyValAspPro 130
Db 295 CTAAATCAGGTCAA-----GAAACAGAAATCAATGACCCCGGAGCT 336
Qy 131 GlnProLeuHisGluArgIleAsnGlnAlaLeuSerSerIle-----ValGlnGln 162
Db 337 AAAAGTGACACACGCCCACTTTGAACAACCTTTATCTCAATACGTCGTACTACAAACGAC 396
Qy 145 GlyAlaLeuValLeuSerAspTyrThrAlalysGlyAlaLeuAlaSer-----ValGlnGln 162
Db 397 GATATTGTTATTGTAGCTGGAGGTACCGAACAGTATTCCAAGTGATGCTTATGCACAA 456
Qy 163 MetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAsp 182
Db 457 ATAGCACAAATCAGTGAATAAAGTGTGCGCAACTAGTTGTGATGCGAGAAAAGATCTT 516
Qy 183 PheGluArg-----TyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPhe 199
Db 517 GTTGAACAGCATATTACCATATCGA---CCATTATTATTATAACCAACAAAGATCAATTA 573
Qy 200 GluAlaValValGly---LysCysLysThrGluGluGluIleValGluArgGlyMetLys 218
Db 574 GAAGTAATGTTTAACACATACAGTCAAGAGTGAAGATGATTAATAATATGGTAAAGAA 633
Qy 219 LeuIleAlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGlnGlyMetSer 238
Db 634 ATTTTGAATAAAGGGCCACATCATCTATCTTTCATTTGGTGGGATGGTGCAATATAT 693
Qy 239 LeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAla-----Gln 254
Db 694 GTTGATCAA-----CATCAAGCATTAATGAAGCTGTGAATCCACAAGGA 735
Qy 255 GluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeu 274
Db 736 CATGATGTTAATACAGTAGGATCTGGTATAGTACAGTGGCAGGTATGATGACGGGTG 795
Qy 275 AlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGlyValVal 294
Db 796 TCCATGGGTCTTAATATAGATGAAGCT-----TTTCAACAGCGCGTAGCTTCA----- 843
Qy 295 ValGlyLysLeuGlyThrSerThr 302
Db 844 -----GGAACAGCGGACT 855
```

RESULT 9

PCT-US02-32727-132

; Sequence 132, Application PC/TUS0232727

; GENERAL INFORMATION:

; APPLICANT: Mitcham, Jennifer

; APPLICANT: Skeiky, Yasir

; APPLICANT: Persing, David

; APPLICANT: Bhatia, Ajay

; APPLICANT: Maisonneuve, Jean Francois

; APPLICANT: Zhang, Yanni

; APPLICANT: Wang, Siqing

; APPLICANT: Jen, Shylan

; APPLICANT: Lodes, Michael

; APPLICANT: Benson, Darin

; APPLICANT: Jones, Robert

; APPLICANT: Carter, Darrick

; APPLICANT: Barth, Brenda

; APPLICANT: Douglass, John

; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V

; FILE REFERENCE: 210121.514C1

; CURRENT APPLICATION NUMBER: PCT/US02/32727

```
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 132
; LENGTH: 13380
; TYPE: DNA
; ORGANISM: Propionibacterium acnes
PCT-US02-32727-132
```

Alignment Scores:

Pred. No.:	0.0449	Length:	13380
Score:	133.00	Matches:	124
Percent Similarity:	34.96%	Conservative:	69
Best Local Similarity:	22.46%	Mismatches:	183
Query Match:	5.56%	Indels:	177
DB:	1	Gaps:	27

US-09-912-020-325 (1-477) x PCT-US02-32727-132 (1-13380)

```
Qy 13 ValMetValValGlyAspValMetLeuAspArgTyr-----TrpIleGlyProThrSer 30
Db 8449 GTGCTCACGATGGGACGAATCGGTGTAGACATCTATCCGCTGCAGTACGGG----- 8499
Qy 31 ArgIleSerProGluAlaProValProValValLysValAsnThrIleGluAlaArgPro 50
Db 8500 -----GTGGGGCTTGAGGATGTCACGCTCGTTCGGAAGTTCCCTT 8538
Qy 51 GlyGlyAla---AlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeu 69
Db 8539 GCGGTAGTCCGACCAATGTGGCGCTGCGGCTGCCAGCTGCGTCATTCTGCCGCGATC 8598
Qy 70 ValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSerLeuAlaAspVal 89
Db 8599 GTCACGGCAGTAGGCGATGATCCCTTCGGCGCGTCTGTGCGCGCGAGATGCCAGACTC 8658
Qy 90 AsnValLysCysAspPheValSerVal-----ProThr----- 100
Db 8659 GGTGTCTATGACAACTACGTGCTGCTCAATTCCGAATTCGCCACCTCCGGTACGTTTGC 8718
Qy 101 -----HisProThrIleThrLys 106
Db 8719 GAGATCTTTCTCCCGACGACTTCCCGCTCTACTTCTACCGCGCTCTTACCCTCCAGAT 8778
Qy 107 LeuArgValLeuSer-----ArgAsnGlnGlnLeuIleArg 118
Db 8779 CTGAGGATCATCTCCGACGACGCTGCTGAGGATGCTGTGAGGAATCGCAAAATCTGTGG 8838
Qy 119 LeuAspPheGluGluGlyPheGluGlyValAspProGlnProLeuHisGluArgIleAsn 138
Db 8839 CTT-----TCGGTGACTGGCTTGTGTCAACACAGCTTCTCTACGACGCACATTGC 8886
Qy 139 GlnAlaLeuSerSerIleGly-----AlaLeuValLeuSerAspTyrAlaLys 154
Db 8887 CGAGCCCTTGAGGTGCGCGGTAAAGGCAGACACACCGCTCTCCGACCTCGACTACCGT--- 8943
Qy 155 GlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLysAlaGlyValProVal 174
Db 8944 -----TCGATGTTCTGGCAGTCCGCGCGGAGGCTCATGAGAGCTCTCAGCGGTG 8994
Qy 175 LeuIleAspProLysGlyThrAspPheGluArgTyrArgGlyAlaThrLeuLeuThrPro 194
Db 8995 TTG-----CCGACAGTGCAGTGGTGGCGATCGGA 9021
Qy 195 AsnLeuSerGluPheGluAlaValValGlyLysCys-----LysThrGluGluGlu 211
Db 9022 AATCCGAGGAGGTGTGAGATGGCGGTGCGGAGTCCGACCGCGGAGCTGCTCAGCGCA 9081
Qy 212 IleValGluArgGlyMetLysLeu-----IleAlaAspTyrGluLeuSerAlaLeuLeu 229
Db 9082 CTGCTTGATCAGGCGGTTTCAGCTTCCCTCGCTCAACAGGAGGCGCCCAAGGTACCTTGCC 9141
Qy 230 ValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMet 249
Db 9142 ATGACCCCGTGAGGAGCGC-----GTGGAGGTGCGCCCGCCCAATC----- 9183
```



```
QY 250 ProThrGlnAlaGlnGluValTyrAspValThr-----GlyAlaGlyAspThrVal 266
Db 9184 -----GACGTTACCAATGGCTCGGAGCGCGAGATTCCTTC 9219

QY 267 IleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePhe 286
Db 9220 GCGCGCAGCTATGTCACGGTTGCTAGAGGATGAGGCTCCCGCAGAGCATTCAGCA 9279

QY 287 AlaAsnAlaAlaGlyValValValGlyLeuGlyThrSerThrValSerProile 306
Db 9280 GCGTCGCGCCAGAGCGCTCGTCAGAGCGGTTGGAGTGTTCACCGCGATGCCCTCG 9339

QY 307 GluLeuAlaAsnAlaValArgGlyArgAlaAspThr-----Gly 319
Db 9340 GAACCGAAGT-----GTTCGGATGATGGCGAACAATCCGGACATTCGCCATCAAGAGGA 9395

QY 320 PheGlyValMetThrGlu----- 325
Db 9396 GCATCTGTCATGACGCGCACCGCGCTATTGTGACAAGCTGTGTAGATCGCGACTTCGCCGAT 9455

QY 326 -----Glu 326
Db 9456 CCGGTGCGTTTCCGATCGCGTCCGCGACGCTCCCGCAGCGAGCTGGAGCCCGCATGAA 9515

QY 327 GluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGluLysValValMetThrAsn 346
Db 9516 CCCTCATGCTGCTCGCGCTGATCTCTGCTCGCGGTGCA----- 9557

QY 347 GlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaArgLysLeu 366
Db 9558 -----CTCGCAGCGGAGGTGCGTCCGATGCTCATGCTGACCGGGAACACCTC 9605

QY 367 GlyAspArgLeuIleValAla-----ValAsnSerAspAlaSerThrLys 381
Db 9606 CTGAGCGCTGCTGACCGCATTAGACGCTCCCGGGTCAATGGTTCCTGGAACTGCC 9665

QY 382 ArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeuGly 401
Db 9666 GACATGTTGCTCCGACCTCGGGAACCTCGAGGCTCTAGATGGCAAG---GTTGATTCGGT 9722

QY 402 AlaLeuGluAlaValAspTrpVal---ValSerPheGluAlaAspThrProGlnArgLeu 420
Db 9723 TCGATGAACGAGGAGGGTTCGCTGGAGCTTCCTTCGAGCTCGAC-----GACCGAATG 9776

QY 421 -----IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLys 436
Db 9777 ACGGGCTACAGCTGCAAGGAATCCTT----- 9803

QY 437 ProGluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeu 456
Db 9804 GACGAGGATCTGACCGGTGGCAAG-----ATGCTGCTG 9836

QY 457 AspPhe-----GluAspGlyCysSerThrThr 465
Db 9837 AGGTTGATCTATCCAGCAGGATGAGGCTTACCAAG 9872

RESULT 10
US-09-134-000C-1139
; ORGANISM: Enterococcus faecalis
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1139
; LENGTH: 432
```

```
TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1139

Alignment Scores:
Pred. No.: 0.000664 Length: 432
Score: 131.00 Matches: 46
Percent Similarity: 50.38% Conservative: 21
Best Local Similarity: 34.59% Mismatches: 58
Query Match: 5.47% Indels: 8
DB: 5 Gaps: 4

US-09-912-020-325 (1-477) x US-09-134-000C-1139 (1-432)

QY 341 LysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeu 360
Db 37 AAAAAAATCTACTTACCGCACATTTGACCTCTCTACACTATGTCATATTAATTTGTTG 96

QY 361 AlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThr 380
Db 97 AAAAAAGCGAACAACAAGGGGACTACCTGATTTGGACTTTCAACAGATGCTTTTAAT 156

QY 381 LysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeu 400
Db 157 TTAGAAAAAAGAAACAAAGT-----TATTTTTCATATGAAAAACGTAACAAATATTA 210

QY 401 GlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrProGlnArgLeu 420
Db 211 GAGGCAATTCGTTATGTTAGTTATTCCT---GAAACTAGTTGGAAACAAAGATT 267

QY 421 -----IleAlaGlyIleLeuProAspLeuValLysGlyGlyAspTyrLysProGlu 438
Db 268 TCCGATATTGCTGAATTTAAATTTGATCTTTAGTCATGGCGCATGATTGGCAAGGGCT 327

QY 439 GluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeuAsnPhe 458
Db 328 TTTGATTTTCTAGAACTTGAACAACACCGCA-----AAGGTTCTTTATTTGCCACGA 378

QY 459 GluAspGlyCysSerThrThrAsnIleIleLysLysIle 471
Db 379 ACGCCAGAGTATCAACCCACACAGATAAAAAAGAATTA 417

RESULT 11
US-10-264-237-1121
; Sequence 1121, Application US/10264237
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1121
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (895)..(895)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (910)..(910)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (919)..(919)
; OTHER INFORMATION: n equals a,t,g, or c
```

```

Db    761  -----ATCATTACCTTAAAGGCGTGTAAGGATGTGTGGTGTGCACAGACA   |||
Qy    243  GlyLysAlaProLeuHisMetProThrGlnAlaGlnValTyrrspValThrGlyAla   |||
Db    806  GAACCTGAGCCAAAGCACATTCACACAGAGAAAAGTCAAGGCTCTGGATACCACGGGTGCT 865
Qy    263  GlyaspThrValIleGlyValLeuAla 271
Db    866  GGTGACAGCTTTGTGGAACTCTGGCC 892

RESULT 12
PCT-US02-32727-16
; Sequence 16, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhalla, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siding
; APPLICANT: Jen, Shylan
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 16
; LENGTH: 61557
; TYPE: DNA
; ORGANISM: Propionibacterium acnes
PCT-US02-32727-16

Alignment Scores:
Pred. No.:      2.78          Length:      61557
Score:         123.00        Matches:     119
Percent Similarity: 34.09%   Conservative: 61
Best Local Similarity: 22.54% Mismatches:    186
Query Match:     5.14%       Indels:      163
DB:              1          Gaps:         25

US-09-912-020-325 (1-477) x PCT-US02-32727-16 (1-61557)
Qy    4 ThrLeuProGluPheGluArgAlaGlyValMetValValGlyAspValMetLeuAspArg 23
Db    32230 TCATATGCCAGTCAGGTCAGGTCGCCGACGATG-----GGCAGCGCATGCTGTTCGCG 3222
Qy    24 TyrTrp---TyrGlyProThrSerArgIleSerProGluAlaProValProValLys 42
Db    32284 CACTGGTGATCGTCTCCGCCGAGCCT-ACGAACCCCGACGTC----- 3233
Qy    43 ValasnThrIleGluGluArgProGlyGlyAlaAlasnValAlametAsnIleAlaSer 62
Db    32325 ATGTCGAGGATTCCTCAGCTCCGAACGCGGC---TCCCACCTCGAGATTCGTGTCGCGCGC 3233
Qy    63 LeuGlyAlaasnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeu 82
Db    32382 CGTGGTGACAAGCGACTCTC-----ATGACACCGTCACTCCGCAATGCC 324
Qy    83 SerIysSerLeuAlaAspValasnValLysCysAspPheValSerValProThrHisPro 102
Db    32427 CAGGAACACTCTGCCCGACACAACTCAAG-----CGGGCCTCTGAC 374
Qy    103 ThrIleThrLysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAsp----- 120
Db    32469 CTTCGTAACCGCTACTCGCGTCTCGAAGAATAATTCAGGAGTCACTCGCGCTTGACCGCGC 325

```

```
Oy 120 ----- 120
Db 32529 CGCTACGATAGAGTCTACGACATTTCCCATATCCAGGACCAATGTCGTGGGTCG 32588
Oy 121 ----- 123
Db 32589 ATGGTGGTCTTCCGAAGAGCGGTATGCCCGCGAGTCTAGTACCGCTGTTTCAGCATCAAG 32648
Oy 124 GlyPheGluGlyValAspPro---GlnProLeuHisGluArgIleAsnGlnAlaLeuSer 142
|||||
Db 32649 GGATTTGAGGGTCCGACGACTTCGCCGCTATGCACGAGGTCTTTCGTCTCGTGGTGGT 32708
Oy 143 SerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGln 162
|||
Db 32709 GCGCTC-----ATCAGGAC-----CGTGATGCATGGCGCTGCTCAA--- 32747
Oy 163 MetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAsp 182
|||||
Db 32748 -----ACTCTGACGGGACGTTGGCTCCCTTTCGATCGCATCCGACTACCGGATCC 32795
Oy 183 PheGluArgTyrArgGlyAlaThrLeuLeu-----ThrProAsnLeu 196
|||||
Db 32796 CCTCGGAATTCGCTACGCCCCCGACCTCATCTGCTCGATGGTGGTGGCCCTCAGGTG 32855
Oy 197 SerGluPheGluAlaValValGlyLysCysLysThrGluGluGluIleValGluArgGly 216
|||
Db 32856 CATGCCGCCACGAGGTCTTGAAGAGTTCGGCCTGGAGGACGAAATCGCGTGTGTGGC 32915
Oy 217 MetLys-----LeuIleAlaAspTyrGluLeuSerAlaLeuVal 230
|||
Db 32916 CTGGCCAAAGCCGTGGAGAGGTGTGGCTTCTCGCAGGAGTAATGGCGGTCAATTTG--- 32972
Oy 231 ThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLysAlaProLeuHis----- 248
|||||
Db 32973 CCGCAACTTCGGAAGGACTCTACTGCTTCAGCGTCTGCGTGTAGGCTCAGCGTTTC 33032
Oy 249 -----MetProThrGlnAlaGlnGluValTyrAsp--- 258
|||
Db 33033 GCTATTACTTTACCGGTCTAAGCGTTTCCAAAGGCCATGTGGTGGAGTCGTTCTCGACGGT 33092
Oy 259 ValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsn 278
|||||
Db 33093 GTGTGGGGCTGGGAGACCCGCGCAACGCTGCTCTCCACTTTGGATCGGTACGG 33152
Oy 279 SerLeuGluGluAlaCysPhe-----PheAlaAsnAlaAla 290
|||||
Db 33153 TCGCTACGCAAGGCCACCGTTGACGACATTCGAGGTGGCCAGGCTTTGGCCGAAATC 33212
Oy 291 AlaGlyValValValGlyLysLeuGlyThrSerThrValSerProIleGluLeuGluAsn 310
|||||
Db 33213 GCTGGCAAGTGTTCGACGCTTG-----TCCGAGGACAAACAGCGGAGCCATCAAC 33266
Oy 311 AlaValArgGlyArgAla-----AspThrGlyPhe----- 320
|||
Db 33267 ACCGCTACTGCGAGGTGTTGTGAGCCCAACGGCCAGCAGCGGAGATACGCGCATAC 33326
Oy 321 -----GlyValMetThrGluGluGluLeuLys----- 329
|||
Db 33327 CCTTGAGCTGTGACACCGACGACGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33386
Oy 330 -----LeuAlaValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 342
|||
Db 33387 ATTTCTGTGTGGGCGTGTGACGCTACCGCTGCTGGAGGACCTTGGCTGTGTGTGTGTGTGT 33446
Oy 343 Val-----MetThrAsnGlyValPheAspIleLeuHisAlaGlyHis 356
|||
Db 33447 GTCGATATCTTCGCCCGCTGTATGTGGGGCCCTCGTGTGACGAAATTCGCGCCCAATAC 33506
Oy 357 ValSerTyrLeuAla-----AsnAlaArgLysLeuGlyAspArgLeuIle 371
|||
Db 33507 ATTGATCGCTAGCGGTGGTTCGATGTCGTAGCGGCATCATGTTGACGCCCTTAGGG 33566
|||
```

```
Oy 372 ValAlaValAsn----- 375
|||||
Db 33567 GTGGCGGTCAATCGCTCGACGAGCGGCGCATCGACCTGCCATCGTCTTCTTGAAGCG 33626
Oy 376 SerAspAlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluGln 395
|||||
Db 33627 TCCGATGAGCAATGTGCGCGCGCAGGAGTCCAGCGCTCGCTCCGCTCGCGTGAACAA 33686
Oy 396 ArgMetIleValLeuGlyAlaLeu 403
|||
Db 33687 GGTGGTCACTGCTCGACGCGGTT 33710

RESULT 13
US-09-134-000C-1809
; Sequence 1809, Application US/09134000C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1809
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1809

Alignment Scores:
Pred. No.: 0.0136 Length: 810
Score: 120.50 Matches: 38
Percent Similarity: 44.59% Conservative: 28
Best Local Similarity: 25.68% Mismatches: 61
Query Match: 5.04% Indels: 21
DB: 5 Gaps: 3

US-09-912-020-325 (1-477) x US-09-134-000C-1809 (1-810)

Oy 184 GluArgTyrArgGlyAlaThr----- 190
|||||
Db 361 GAGCGCGCGTGGCGGCGACGCTCAAAATCCCGCACCTGGCTTAGAACAAAGTTCTCTGAAGAA 420
Oy 191 -----LeuLeuThrProAsnLeuSerGluPheGluAlaValValGly 204
|||||
Db 421 TTACTAACGTGACAGATATGTTGTACCAATGAACAGAAACCGGAATTTTACAGGC 480
Oy 205 LysCysLysThrGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGlu 224
|||||
Db 481 ATTAATACACAGATGAGCGAGTATGCGTAAGCCGCGAGACACTTCATCAATTAGGG 540
Oy 225 LeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLys 244
|||
Db 541 ATTGAAGCAGTAATTATTACAGTAGTAGTAGTAAGCGCTTTTATGACGTCAATGAGCA 600
Oy 245 AlaProLeuHisMetProThrGlnAlaGlnValTyrAspValThrGlyAlaGlyAsp 264
|||
Db 601 AGTGGTATT---GTGCTCTTTTAAAGTGAAGCGGTGTATACAAACGGCTGCTGGCGAT 657
Oy 265 ThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGly---AsnSerLeuGluGluAla 283
|||
Db 658 ACTTTTATGGCGCATTAAGTAGTATATTAGAAAAGATTTTAGCAATTTGGAGAAGCT 717
Oy 284 CysPhePheAlaAsnAlaAlaAlaGlyValValValGlyLysLeuGlyThrSerThrVal 303
|||
Db 718 ATTCGTTATGGAACAACAGCGCTCTTCGTTGACTGTTCACAGCTTTTGGAGCCCAACCTTCG 777
Oy 304 SerProIleGluLeuGluAlaAsnAla 311
|||
Db 778 ATTCCTTATCAACACGAATGGCA 801
|||
```

```
RESULT 14
US-10-264-213-100/c
; Sequence 100, Application US/10264213
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Lubbers, Mark William
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; FILE REFERENCE: 11000.1043c3
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 12023
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-100

Alignment Scores:
Pred. No.: 0.572 Length: 12023
Score: 120.00 Matches: 85
Percent Similarity: 32.71% Conservative: 55
Best Local Similarity: 19.86% Mismatches: 184
Query Match: 5.01% Indels: 104
DB: 13 Gaps: 13

US-09-912-020-325 (1-477) x US-10-264-213-100 (1-12023)

QY 10 ArgAlaGlyValMetValValGlyAspValMetLeuAspArgTyrTriPtyrGlyProThr 29
Db 7940 AGGGGAACATTCAAAATGTCAAAAATTCGATTGCTATGTCGAGCAGCAGT 7881

QY 30 -SerArgIleSerProGluAlaProValProValValLysValAsnThrIleGluGluAr 49
Db 7880 TCATTAAATGCGGTGATATAATCGACCTTTAGAGAACTAAGAGCTTTGCAAAATT 7821

QY 49 gProGlyGlyAla---AlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaAr 68
Db 7820 CGTTGCGGATCCCGGCAAAATATTGCAATCGGTCAGCAAAATTAGGCCAAAAAGTAGG 7761

QY 68 gLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSerLeuAlaAs 88
Db 7760 ATTATCGGGAAGGTGTCGATGATCAGCTGGTCTATTATGTCAGCAATATACATGCAGA 7701

QY 88 pValAsn-----ValLysCysAsp----- 94
Db 7700 CGTTGCGATTGACACACGCAAAATGTCAGGATGATGCGGGCCACAAAATTTGGCTTAAC 7641

QY 95 -PheValSerVal-----ProThrHisProThrIleThrLysLeuArgValLeuSerAr 112
Db 7640 GTTTACTGAATCATATTAGTCCGAAAGAGCGATATTGATGATATCGAAATGAAGCCGC 7581

QY 112 gAsnGlnGlnLeuIleArgLeuAspPheGluGlyPhe----- 125
Db 7580 CGATTATATCTGACACCGCTGATGTTTCAGAAAGATATTAGCGCAAGCCAAAGATGTT 7521

QY 126 -----GluGlyValAspProGlnProLeuHisGluArgIleAsnGlnAlaLe 141
Db 7520 GGTGATTTCCGGGACCGGATTAGCCAAAGCCGTCGCGAGAGCGGATTTTAAAGGCACT 7461

QY 141 uSerSerIleGlyAlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAspValGl 161
Db 7460 CACA----- 7457

QY 161 nGlnMetIleGlnLeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyTh 181
Db 7456 -----GTGCAAGACTACTTGGGTTGAAGTGGTGTGAA-----CT 7419

QY 181 rAspPheGluArgTyr----- 186

Db 7418 GGATTATGCGCGTATACCTGAAAAATGCCGAGAAACGAGCTCTTTATTATCAATTAGT 7359
QY 187 ----ArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGly-- 204
Db 7358 TGCCGAGCGTGCAGATGATATTGTCGACGCGTGACGAATTTGATGTTCTCGAAAAATCA 7299
QY 205 -LysCysLysThrGluGluIleValGluArgGlyMetLysLeuIleAlaAspTyrGl 224
Db 7298 TCACGGTAAACAGCAGCAGCAATTCGACACTTTTCAAA-----TATGCA 7251
QY 224 uLeuSerAlaLeuValThrArgSerGluGlnGlyMetSerLeuLeuGlnProGlyLy 244
Db 7250 TCCTAAGCTAATTGTCTCAAAAAGCGGGTTCGAAGGTCAAAATCGTATACAAAAGCCGG 7191
QY 244 sAlaProLeuHisMetProThrGlnAlaGlnValValTyrAspValThrGlyValaGlyAs 264
Db 7190 CGATCATTTATCTTTGGAGTTTCAAGACGAAGGTTCCTTAAGAGTTTCGGCGCTGGCGA 7131
QY 264 pThrValIleGlyValLeuAlaAlaThrLeuAlaGlyAsuSerLeuGluGluAlaLac 284
Db 7130 TTCATTGCTGCGGGTTCCTATATGCTTACAGCCAGCAGCTTGGGAATAGAACTGCCTT 7071
QY 284 sPhePheAlaAsnAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSe 304
Db 7070 GAAGTATGCGCTGTCAGCGCTCTATTGATTAAGTCTTTCAGAAAGCGAT 7011
QY 304 rPro-Ile-----GluLeuGluAsnAlaValArgGlyArgAlaAspThrG 319
Db 7010 GCCGATTTGGCGAAACTACGGCATTTATTCACGAAGCAGACAGCAGGAGGTGCATTA 6951
QY 319 lPheGlyValMet-----ThrGluGluG 327
Db 6950 GGATGAGCGCATGAGTTGAAGTCCAGAGCAGCAGCAGCAGCTCAGCCGCGGAGATA 5891
QY 327 lLeuLysLeuAlaValAlaAlaAlaArgLysArg----- 338
Db 6890 CAGTTTCGATTGACACCGCGCAGCGCTCGTCGCTTTTAAACCAACAATATCTTGATG 6831
QY 339 -----GlyGluLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHis 357
Db 6830 TTGATGGTGAGTGCACCCCGTTTGTGCAAGGATCTTTGCAATTTTGGCCATGGCAATG 6771
QY 357 alSerTyrLeuAlaAsnAla 363
Db 6770 TACTGGTTGGGGGAAGCC 6751

RESULT 15
US-09-724-676-45787
; Sequence 45787, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45787
; LENGTH: 1877
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-45787

Alignment Scores:
Pred. No.: 0.0519 Length: 1877
Score: 119.50 Matches: 72
Percent Similarity: 35.67% Conservative: 40
Best Local Similarity: 22.93% Mismatches: 127
Query Match: 4.99% Indels: 75
DB: 5 Gaps: 13

US-09-912-020-325 (1-477) x US-09-724-676-45787 (1-1877)
```

```
QY 146 AlaLeuValLeuSerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGln 165
   |||   ::|   |||||   ::|   |||   ::|
Db 135 GCACGGCCCATGGTGACTAC-----CTCATCGTAGCGTCGCACACCGATGAGGAG 185

QY 166 LeuAlaArgLysAlaGlyValProValLeuIleAspProLysGlyThrAspPheGluArg 185
   ::|::|::|   |||   |||||   |||   |||   |||
Db 186 ATGCCAAGCACAAAGGGCCCGCGTGTTC-----ACTCAGGAGGAGAGA 230

QY 186 TyrArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValValGlyLys 205
   |||::|::|
Db 231 TACAAG-----ATGGTGCAGGCC 248

QY 206 CysLysThrGluGluIleVal-----GluArgGly 216
   |||   ::|::|::|::|   |||
Db 249 ATCAATGGGTGACGAGGTGGTCCAGCGGCTCCCTACCTACCTACCTAGACACCCCTG 308

QY 217 MetLysLeuIleAlaAspTyrGluLeuSerAlaLeuLeuValThrArgSerGluGlnGly 236
   |||   |||::|::|   ::|   ::|::|   |||
Db 309 GACAAATACAACTGTGACTTCTGTTCACGGCAATCACATCACCTGACTGTAGATGGC 368

QY 237 Met-----SerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGln 252
   |||   ::|   |||   |||::|   |||::|
Db 369 CGGACACCTTATGAGCAAGTAAAGCAGGTGGAGGTACAGAGAATGCAAGCCGACGCAA 428

QY 253 AlaGlnGluValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAla 272
   |||::|   |||   |||   ::|::|   |||
Db 429 GGGGTGTCCACCACACACCTCTGTGGCCGC-----ATGCTGTGGTA 470

QY 273 ThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaGly 292
   |||   ::|::|   |||   ::|   |||
Db 471 ACCAAAGCCCATCACAGCAGCAGAGATGTCTCTGAGTACCGGGAGTATGCAGACAGT 530

QY 293 ValValValGlyLysLeuGlyThrSerThrValSerProIleGluLeuGluAsnAlaVal 312
   |||::|
Db 531 -----TTTGGCAAG-----TGCCCT 545

QY 313 ArgGlyArgAlaAsp---ThrGlyPheGlyValMetThrGluGluLeuLysLeuAla 331
   |||||   |||||   |||   ::|
Db 546 GTTGGCGGGAACCCCTGGACCGGGGTATCCAGTTCTCTGCAGACATCTCAGAGATCATC 605

QY 332 ValAlaAlaAlaArgLysArg-----GlyGluLysValValMetThrAsnGlyVal 348
   |||::|   |||   |||||   |||::|   |||
Db 606 CAGTTTCTCTCTGGAAGGAGGCCGCCAGCCAGGGGAGACACTCATCTATGTGGTGGTGC 665

QY 349 PheAspIleLeuHisAlaGlyHisValSerTyrIleuAlaAsnAlaArgLysLeuGlyAsp 368
   ||||::|   |||   |||||   |||::|   |||
Db 666 TTCGACCTGTTCACATCGGGCATGTGGACTTCTCTGGAGAAGGTGCACAGGCTGGCAGAG 725

QY 369 Arg-----LeuIleValAlaValAsnSerAspAlaSerThrLysArgLeuLysGlyAsp 386
   |||   ::|::|   ::|::|   |||   |||||
Db 726 AGGCCCTACATATCGGGCGCTTACACTTTGACCAGGAGTCAATCTACAAAGGGGAAG 785

QY 387 SerArgProValAsnProLeuGluGlnArgMet----- 397
   ::|   |||::|   |||   ::|
Db 786 AACTACCCCATCATGAATCTGCATGAACGACTCTGAGCGTGTGGCTGCCCGGTACGTG 845

QY 398 -----IleValLeuGlyAlaLeuGluAlaVal-----AspTrpValValSerPheGlu 413
   ::|::|::|   |||   |||||   |||   ::|
Db 846 TCAGAAAGTGTGTATGGAGCCCGCTACCGCGGTCTACAGCAGAGAGCTCCTAAGTCTCAAG 905

QY 414 GluAspThrProGlnArgLeuIleAlaGlyIleLeuProAsp 427
   |||
Db 906 GTGGACCTGTGTCTACGGCAAGACAGAAATTTATCCCTGAC 947
```

Search completed: November 27, 2002, 02:55:41
Job time : 165 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 24, 2002, 23:18:35 ; Search time 3218 Seconds
(without alignments)
4313.871 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERAGVMVGVDM.....FEDGCTNTIIKKIQDQKKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 1567388

Minimum DB seq length: 0
Maximum DB seq length: 300
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/Cyg2_I/USPTO_spool/US09912020/runat_20112002-150259-11086/app-query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2nszlm300.rge -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=300
-USER=US09912020_8CGN_1_1_3637_@runat_20112002-150259-11086 -NCPU=6 -ICPU=3
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hug:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	263	11.0	294	6	AR203747	AR203747 Sequence
2	170	7.1	243	6	AX311342	AX311342 Sequence
3	153.5	6.4	273	1	AF358705	AF358705 Helicobac
C 4	120	5.0	272	6	AX188888	AX188888 Sequence
C 5	112	4.7	292	6	AX401588	AX401588 Sequence
6	73	3.1	264	1	MG001777	U01777 Mycoplasma
7	72	3.0	300	3	AF186316	AF186316 Echinomet
8	72	3.0	300	3	AF186317	AF186317 Echinomet
9	71	3.0	300	3	ESU39543	U39543 Echinometra
C 10	70	2.9	245	6	AX188889	AX188889 Sequence
11	70	2.9	300	3	AF186281	AF186281 Echinomet
12	69.5	2.9	285	3	AF186277	AF186277 Echinomet
13	69.5	2.9	285	3	AF186279	AF186279 Echinomet
14	69.5	2.9	285	3	AF186282	AF186282 Echinomet
15	69.5	2.9	285	3	AF186284	AF186284 Echinomet
16	69.5	2.9	285	3	AF186285	AF186285 Echinomet
17	69.5	2.9	285	3	AF186288	AF186288 Echinomet
18	69.5	2.9	285	3	AF186297	AF186297 Echinomet
19	69.5	2.9	285	3	AF186305	AF186305 Echinomet
20	69.5	2.9	285	3	AF186306	AF186306 Echinomet
21	69.5	2.9	285	3	AF186308	AF186308 Echinomet
22	69.5	2.9	285	3	AF186309	AF186309 Echinomet
23	69.5	2.9	285	3	AF186310	AF186310 Echinomet
24	69.5	2.9	285	3	AF186321	AF186321 Echinomet
25	69.5	2.9	285	3	AF186322	AF186322 Echinomet
26	69.5	2.9	285	3	AF186324	AF186324 Echinomet
27	69.5	2.9	285	3	AF186331	AF186331 Echinomet
28	69.5	2.9	285	3	AF186337	AF186337 Echinomet
29	69.5	2.9	285	3	AF186338	AF186338 Echinomet
30	69.5	2.9	285	3	AF186339	AF186339 Echinomet
31	69.5	2.9	285	3	AF186344	AF186344 Echinomet
32	69.5	2.9	285	3	AF186347	AF186347 Echinomet
33	69	2.9	288	3	AF186307	AF186307 Echinomet
34	67	2.8	208	10	AF196854	AF196854 Mus muscu
C 35	67	2.8	270	13	USU42871	U42871 Unidentifie
36	66.5	2.8	287	14	AF457273	AF457273 HIV-1 G23
37	65.5	2.7	271	9	HA301468	AJ301468 Homo sapi
38	65.5	2.7	280	6	AR128860	AR128860 Sequence
39	65.5	2.7	285	3	AF186328	AF186328 Echinomet
40	65.5	2.7	285	3	AF186343	AF186343 Echinomet
41	64	2.7	237	6	AX119171	AX119171 Sequence
42	64	2.7	275	6	AX341713	AX341713 Sequence
C 43	64	2.7	288	11	G06129	G06129 human STS W
44	64	2.7	297	14	AY047445	AY047445 HIV-1 iso
45	63.5	2.7	285	3	AF186349	AF186349 Echinomet

ALIGNMENTS

RESULT 1

AR203747
LOCUS AR203747 294 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 186 from patent US 6365401.
ACCESSION AR203747
VERSION AR203747.1 GI:21500216
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 294)
AUTHORS Mahan,M.J., Comer,C.P. and Heithoff,D.M.
TITLE Method and probes for the identification of microbial genes specifically induced during host infection
JOURNAL Patent: US 6365401-A 186 02-APR-2002;
FEATURES
Location/Qualifiers
source 1..294
BASE COUNT 67 a 73 c 96 g 58 t
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-08 Length: 294
Score: 263.00 Matches: 72
Percent Similarity: 79.59% Conservative: 6
Best Local Similarity: 73.47% Mismatches: 12
Query Match: 10.99% Indels: 10
DB: 6 Gaps: 5
US-09-912-020-325 (1-477) x AR203747 (1-294)
QY 267 IleGlyValLeuAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhePhe 286
|||||
Db 2 ATCGCGTGTGGCGGACCTG-GCCGCGGAAATACCTGGAAGAGGCGTGTATTTC 60
QY 287 AlaAsnAlaAlaGlyValValGlyLeuGlyThrSerThrValSerProIle 306
|||||
Db 61 GCCAATCGCGCGGCGGAGTGGTAGGTAAACTCGGACGTCACCGGTTCCCTATT 120
QY 307 GluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGlu 326
|||||
Db 121 GAGCTGGAACACGAGTGGCGGAGC- -GATACCGGCTTCGGCTTATGACCGAGAG 176
QY 327 GluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGlyValValMetThrAsn 346
|||||
Db 177 GAGTTGAGACAGCGCGTGGCGGCGGTAAGTC- -CGGAGAAGTGTGATGACCAAC 230
QY 347 GlyValPheAspIle- -LeuHis- - -AlaGlyHisValSerTyrLeu 360
|||||
Db 231 - -CGGTTGATATCTGACGGCATTTATGCGCAACTGACCTATCGGATACTTA 281
RESULT 2
AX311342
LOCUS AX311342 243 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 4327 from Patent WO0190366.
ACCESSION AX311342
VERSION AX311342.1 GI:17896901
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Leach,M.D. and Shimkets,R.A.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0190366-A 4327 29-NOV-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source 1..243
BASE COUNT 51 a 68 c 83 g 41 t
ORIGIN

Alignment Scores:
Pred. No.: 0.00563 Length: 243
Score: 170.00 Matches: 32
Percent Similarity: 65.82% Conservative: 20
Best Local Similarity: 40.51% Mismatches: 27
Query Match: 7.10% Indels: 0
DB: 6 Gaps: 0
US-09-912-020-325 (1-477) x AX311342 (1-243)
QY 378 AlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluInArgMet 397
|||||
Db 1 GCCTCGTGAAGGCTCTGAAAGGCGCCACCGCCGCGTGCAGGACGAGACGCCGAGCC 60
QY 398 IleValLeuGlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrPro 417
|||||
Db 61 GCCGTTATGCGGAGCATTAAGGCGGTGGCGATGTCACGTTTTCGACGAGATACGCCG 120
QY 418 GlnArgLeuIleAlaGlyIleLeuProAspLeuValLysGlyLysTyrLysPro 437
|||||
Db 121 ATCGAGCTCCTCGCGCGCTCTCGCGGAGTGTATCGTCAAGGAGCGGACTATCGAGAA 180
QY 438 GluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeu 456
|||||
Db 181 GACCAAGTGTGTCGCGCGGACCTTGTGAAGAAACACGAGGAGGCTGCTACTTGTG 237
RESULT 3
AF358705
LOCUS AF358705 273 bp DNA linear BCT 17-OCT-2001
DEFINITION Helicobacter hepaticus putative ADP-D-glycero-D-mannoheptose synthase (rfae) gene, partial cds.
ACCESSION AF358705
VERSION AF358705.1 GI:16265977
KEYWORDS Helicobacter hepaticus.
SOURCE Helicobacter hepaticus.
ORGANISM Helicobacter hepaticus
REFERENCE 1 (bases 1 to 273)
AUTHORS Ge,Z., Feng,Y. and Fox,J.G.
TITLE Helicobacter hepaticus genome: construction of an ordered cosmid library and sequence analysis of the selected genomic regions
JOURNAL Unpublished
AUTHORS Ge,Z., Feng,Y. and Fox,J.G.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2001) Division of Comparative Medicine, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
FEATURES
Location/Qualifiers
source 1..273
/organism="Helicobacter hepaticus"
/strain="3B1; ATCC 51449"
/db_xref="ATCC:51449"
/db_xref="taxon:32025"
gene <1..>273
/gene="rfae"
CDS <1..>273
/gene="rfae"
/transl_table=1
/product="putative ADP-D-glycero-D-mannoheptose synthase"
/protein_id="AAL16695.1"
/db_xref="GI:16265978"
/translation="PEAPVQVVDVVDENNRLLGACNVNHLNALNQVFCVGVGND EAGFWLGEKLESMGVDISYLFVDTSRPTTKTRVITIANQOVLVRDRES"
BASE COUNT 75 a 41 c 71 g 86 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0669 Length: 273
Score: 153.50 Matches: 37
Percent Similarity: 55.56% Conservative: 13

[illegible]

```

source
1. .300
/organism="Echinometa mathaei"
/isolate="Flc.10"
/db_xref="taxon:31178"
<1. .>300
/codon_start=1
/product="bindin precursor"
/protein_id="AAF07131.1"
/db_xref="GI:6319068"
/translation="IGNYPQARNPPMGNGNYPVPGQAPMGQLAQOQYAAPGMGGPVGG
GGGAWAGFGGAGVGGGAGPPEFGMPAEAGEEDKDYSSVDEETISAK"
mat_peptide
1. .>300
/product="bindin"
BASE COUNT 71 a 69 c 110 g 50 t
ORIGIN
Alignment Scores:
Pred. No.: 7.63e+03 Length: 300
Score: 72.00 Matches: 33
Percent Similarity: 40.19% Conservative: 10
Best Local Similarity: 30.84% Mismatches: 31
Query Match: 3.01% Indels: 33
DB: 3 Gaps: 4
US-09-912-020-325 (1-477) x AF186317 (1-300)
Qy 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValThrAspValThrGly 261
Db 58 CCGGTCAGCACCT-----ATGGTCAGCTAGCCCAACAGGTTACCTGCTCTGGA 111
Qy 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 112 ATGGGTGGACCGTCGGCGT-----ATGGGTGGACCGTCGGCGT----- 132
Qy 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSe 301
Db 133 -----GGTGGTGGTCGAATGGCCGGCCCTTCGGTGGGGAGCTGGCGCTGTC 180
Qy 301 rThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheG1 321
Db 181 GGTGGTGGGGAGCTGGACCTCAGAAATTTGGAGAGA----- 217
Qy 321 yValMetThrGluGluGluLeuLys-----LeuAlaValAlaAlaAlaAr 336
Db 218 -TGCCCGAAGCGAAGGAGCTGAAGAGGGGTGATAAGGACTACAGTAGCGTCGATGAG 276
Qy 336 gLysArgGlyGluLysVal 342
Db 277 GAAGAGACAACAATTAGT 295
RESULT 9
ESU39543 300 bp DNA linear INV 21-MAY-1996
LOCUS Echinometa sp. clone A-B10 bindin gene, partial cds.
DEFINITION U39543
ACCESSION U39543.1 GI:1122528
VERSION U39543.1
KEYWORDS Echinometa sp.
SOURCE Echinometa sp.
ORGANISM Echinometa sp.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinomtridae;
Echinometa.
1 (bases 1 to 300)
Metz, E.C. and Palumbi, S.R.
Positive selection and sequence rearrangements generate extensive
polymorphism in the gamete recognition protein bindin
Mol. Biol. Evol. 13 (2), 397-406 (1996)
96164534
MEDLINE
PUBMED 8587504
REFERENCE
2 (bases 1 to 300)
Metz, E.C. and Palumbi, S.R.
Direct Submission
TITLE Submitted (26-OCT-1995) Edward C. Metz, Zoology, University of

```

```

Hawaii, Honolulu, HI 96822, USA
Location/Qualifiers
1. .300
/organism="Echinometa sp."
/strain="type A"
/db_xref="taxon:43350"
/clone="A-B10"
<1. .>300
/notes="sperm acrosomal protein"
/product_start=1
/product="bindin"
/protein_id="AAB00431.1"
/db_xref="GI:1122529"
/translation="VGNYPQARNADPGNYPVPGCAPMGQLAQOQYAAPGMGGPVGG
GGGARAGPIGGAGPVGGGAGPPEFGMPAEAGEEGDEDEYSIAEETISAK"
BASE COUNT 70 a 69 c 111 g 50 t
ORIGIN
Alignment Scores:
Pred. No.: 8.79e+03 Length: 300
Score: 71.00 Matches: 33
Percent Similarity: 40.19% Conservative: 10
Best Local Similarity: 30.84% Mismatches: 31
Query Match: 2.97% Indels: 33
DB: 3 Gaps: 4
US-09-912-020-325 (1-477) x ESU39543 (1-300)
Qy 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValThrAspValThrGly 261
Db 58 CCGGTCAGCACCT-----ATGGTCAGCTAGCCCAACAGGTTACCTGCTCTGGA 111
Qy 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 112 ATGGGTGGACCGTCGGCGT-----ATGGGTGGACCGTCGGCGT----- 132
Qy 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSe 301
Db 133 -----GGTGGTGGTCGAAGGCGCGGCTATCGTGGGGAGCTGGCGCTGTC 180
Qy 301 rThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheG1 321
Db 181 GGTAGTGGGGAGCTGGACCTCCAA-----ACTTTGAGGG 216
Qy 321 yValMetThrGluGluGluLeuLys-----LeuAlaValAlaAlaAlaAr 336
Db 217 ATGCTCGAAACCGGAGGAGCTGAAGGGAGGTGATGAGACTACAGTAGCATCGCTGAG 276
Qy 336 gLysArgGlyGluLysVal 342
Db 277 GAAGAGACAACAATTAGT 295
RESULT 10
AX188889/c
LOCUS AX188889
DEFINITION Sequence 90 from Patent WO0148209.
ACCESSION AX188889
VERSION AX188889.1 GI:15142430
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 245)
Forsyth, R.A., Ohlsen, K.L. and Zyskind, J.W.
Genes identified as required for proliferation of E. coli
Patent: WO 0148209-A 90 05-JUL-2001;
Elitra Pharmaceuticals, Inc. (US)
Location/Qualifiers
1. .245
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 57 a 53 c 55 g 80 t

```



```

Db 58 CCCGTCAGCACCT-----ATGGTTCAGCTAGCCCCAACAGGTTACGCTGCTCCTCGGA 111
Qy 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 112 ATGGTGGACCGCTCGGTGGT----- 132
Qy 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValVal-ValGlyLysLeuGlyThrSe 301
Db 133 -----GTTGGTGGAGCAATGCGCGGCTATCGTGGGGAGCTGGGCTGTC 180
Qy 301 rThrValSerProIleGluLeuGluAsnAlaValIleGlyArgAlaAspThrGlyPheG1 321
Db 181 GGTGGTGGGGAGCTGGGCCCTCCAGATTTCGAGAGA----- 217
Qy 321 yValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGluGly 341
Db 218 -TGCCCGAAGAGGAGGTGATAAGGACTACAGTAGCTGATGAGGAAGACACACAATT 276
Qy 341 sVal 342
Db 277 AGTG 280

RESULT 13
LOCUS AF186279 285 bp DNA linear INV 10-NOV-1999
DEFINITION Echinomatra mathaei isolate Guam4.1 bindin precursor, gene, partial
cds.
ACCESSION AF186279
VERSION AF186279.1 GI:6318945
KEYWORDS Echinomatra mathaei.
SOURCE Echinomatra mathaei.
ORGANISM Echinomatra mathaei.
REFERENCE 1 (bases 1 to 285)
AUTHORS Palumbi, S.R.
TITLE All males are not created equal: fertility differences depend on
gamete recognition polymorphisms in sea urchins
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (22), 12632-12637 (1999)
MEDLINE 20006286
PUBMED 10535974
REFERENCE 2 (bases 1 to 285)
AUTHORS Palumbi, S.R.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1999) Organismic & Evolutionary Biology, Harvard
University, 16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES
source
location/Qualifiers
1..285
/organism="Echinomatra mathaei"
/isolate="Guam4.1"
/db_xref="taxon:31178"
<1..>285
/codon_start=1
/product="bindin precursor"
/protein_id="AA07093.1"
/db_xref="GI:6319030"
/translation="YGNYPQARNPPMGGGNYVPQGPQMGLAQOQGYAAPCMGPGVGG
mat_peptide 1..>285
BASE COUNT 66 a 66 c 104 g 49 t
ORIGIN
Alignment Scores: 1.02e+04 Length: 285
Pred. No.: 69.50 Matches: 30
Percent Similarity: 39.22% Conservative: 10
Best Local Similarity: 29.41% Mismatches: 34
Query Match: 2.90% Indels: 28
DB: 3
US-09-912-020-325 (1-477) x AF186279 (1-285)

```

```

Qy 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyAspValThrGly 261
Db 58 CCCGTCAGCACCT-----ATGGTTCAGCTAGCCCCAACAGGTTACGCTGCTCCTCGGA 111
Qy 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu 281
Db 112 ATGGTGGACCGCTCGGTGGT----- 132
Qy 282 GluAlaCysPhePheAlaAsnAlaAlaAlaGlyValVal-ValGlyLysLeuGlyThrSe 301
Db 133 -----GTTGGTGGAGCAATGCGCGGCTATCGTGGGGAGCTGGGCTGTC 180
Qy 301 rThrValSerProIleGluLeuGluAsnAlaValIleGlyArgAlaAspThrGlyPheG1 321
Db 181 GGTGGTGGGGAGCTGGGCCCTCCAGATTTCGAGAGA----- 217
Qy 321 yValMetThrGluGluGluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGluGly 341
Db 218 -TGCCCGAAGAGGAGGTGATAAGGACTACAGTAGCTGATGAGGAAGACACACAATT 276
Qy 341 sVal 342
Db 277 AGTG 280

RESULT 14
LOCUS AF186282 285 bp DNA linear INV 10-NOV-1999
DEFINITION Echinomatra mathaei isolate Guam5.2 bindin precursor, gene, partial
cds.
ACCESSION AF186282
VERSION AF186282.1 GI:6318948
KEYWORDS Echinomatra mathaei.
SOURCE Echinomatra mathaei.
ORGANISM Echinomatra mathaei.
REFERENCE 1 (bases 1 to 285)
AUTHORS Palumbi, S.R.
TITLE All males are not created equal: fertility differences depend on
gamete recognition polymorphisms in sea urchins
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (22), 12632-12637 (1999)
MEDLINE 20006286
PUBMED 10535974
REFERENCE 2 (bases 1 to 285)
AUTHORS Palumbi, S.R.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1999) Organismic & Evolutionary Biology, Harvard
University, 16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES
source
location/Qualifiers
1..285
/organism="Echinomatra mathaei"
/isolate="Guam5.2"
/db_xref="taxon:31178"
<1..>285
/codon_start=1
/product="bindin precursor"
/protein_id="AA07096.1"
/db_xref="GI:6319033"
/translation="YGNYPQARNPPMGGGNYVPQGPQMGLAQOQGYAAPCMGPGVGG
mat_peptide 1..>285
BASE COUNT 67 a 67 c 103 g 48 t
ORIGIN
Alignment Scores: 1.02e+04 Length: 285
Pred. No.: 69.50 Matches: 30
Percent Similarity: 39.22% Conservative: 10
Best Local Similarity: 29.41% Mismatches: 34
Query Match: 2.90% Indels: 28

```

DB:	3	Gaps:	3
US-09-912-020-325 (1-477) x AF186282 (1-285)			
QY 242 ProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspValThrGly	261		
Db 58 CCGGTCAGCACCT-----ATGGGTACAGTACCCCAACAGGTTACGCTGCTCCCGGA	111		
QY 262 AlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGlu	281		
Db 112 ATGGTGGTACCGGTCGGTGGT-----	132		
QY 282 GluAlaCysPhePheAlaAsnAlaAlaGlyValVal-ValGlyLysLeuclyThrSe	301		
Db 133 -----GGTGGTGGTCAATGCGCGGCCCTACGTTGGGGAGCTGGGCTGTC	180		
QY 301 rThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheG1	321		
Db 181 GGTGGTGGGGAGCTGGGCTCCAGAAATTGGAGACA-----	217		
QY 321 yValMetThrGluGluLeuLysLeuAlaAlaAlaAlaAlaAArgLysArgGlyGluLy	341		
Db 218 -TGCCCGAAGAGGAGGTGATAGGACTACAGTAGCTCGATGAGGAAGACAACAATT	276		
QY 341 sVal 342			
Db 277 AGTG 280			
RESULT 15			
AF186284			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
Source			
CDS			
mat_peptide			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 24, 2002, 23:16:44 ; Search time 294 Seconds
(without alignments)
3653.752 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLEPERAGVMVGVDM.....FEDGGSTTNIIKKIQDKKG 477

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 2819718

Minimum DB seq length: 0
Maximum DB seq length: 300

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frames_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09912020/runat_20112002_150259_11075/app_query.fasta_1.647
-DB=N_Geneseq.101002 -QFMT=fastap -SUFFIX=p2nslm300.rng -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=300 -USER=US09912020_8CGN_1_1_281 -runat_20112002_150259_11075 -NCPUP=6
-ICPU=3 -NO_XLPHY -NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	170	7.1	243	24	ABN77217
2	162	7.1	249	18	AA767445
3	120	5.0	272	22	AAH81290
4	112	4.7	292	24	ABK63357
5	107.5	4.5	285	24	ABL75687
6	84.5	3.5	300	21	AAA67292
7	71.5	3.0	300	20	AAZ14636
8	70	2.9	245	22	AAH81291
9	67.5	2.8	287	24	ABL74436
10	66.5	2.8	201	21	AA67278
11	65.5	2.7	279	21	ABG63223
12	64	2.7	237	22	AAH52260
13	64	2.7	275	24	ABL38371
14	64	2.7	298	24	ABL76194
15	63.5	2.7	267	22	AAK17810
16	63.5	2.7	267	22	AAK43661
17	63.5	2.7	267	22	AAI49702
18	63.5	2.7	267	22	AAI09966
19	63.5	2.7	284	23	AA586615
20	63	2.6	222	24	ABN78334
21	62	2.6	149	21	AAAL4523
22	62	2.6	149	21	AAAL4524
23	62	2.6	149	22	AA502280
24	62	2.6	149	22	AA502281
25	62	2.6	245	22	ABA88721
26	62	2.6	252	21	AAAL4575
27	62	2.6	258	21	AAAL4570
28	62	2.6	283	22	ABA12944
29	61.5	2.6	225	15	AAQ57460
30	61.5	2.6	299	19	AAV20211
31	61	2.5	176	24	ABL63015
32	61	2.5	264	16	AA26654
33	61	2.5	294	18	AA763569
34	60.5	2.5	234	24	ABR80352
35	59.5	2.5	248	14	ABR81018
36	59.5	2.5	268	15	AAQ57764
37	59.5	2.5	286	20	AAV87597
38	59.5	2.5	294	23	AA554270
39	59	2.5	117	22	ABA71890
40	59	2.5	117	22	AAK20273
41	59	2.5	117	22	AAK46346
42	59	2.5	117	22	AAI52242
43	59	2.5	117	22	ABS20664
44	59	2.5	197	19	AAV31991
45	59	2.5	197	20	AAV55580

ALIGNMENTS

RESULT 1

ABN77217

ID ABN77217 standard; cDNA; 243 BP.

AC ABN77217;

DT 08-JUL-2002 (first entry)

DE Human ORF2164 cDNA, SEQ ID NO:4327.

Human; ORF; open reading frame; ORF; drug screening; diagnosis;
disease monitoring; cytokine; cell proliferation; cell differentiation;
immune modulation; haematopoiesis regulation; tissue growth;
angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;
thrombolytic; tumour inhibition; bodily characteristic; fertility;
behaviour; cancer; proliferative disorder; neurological disorder;
cardiovascular disease; immune system disorder; organ transplantation;
tissue growth disorder; tissue regeneration disorder; diabetes mellitus;

KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiac; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200190366-A2.
 PN
 XX 29-NOV-2001.
 PD
 XX
 PF 24-MAY-2001; 2001WO-US17076.
 PR
 XX 24-MAY-2000; 2000US-206690P.
 XX (CURA-) CURAGEN CORP.
 PA
 XX Leach MD, Shmkets RA;
 PI
 XX WPI; 2002-106200/14.
 DR P-PSDB; ABP33191.
 DR
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation
 XX
 PS Claim 1; Page 1338; 2508pp; English.
 CC
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX
 SQ Sequence 243 BP; 51 A; 68 C; 83 G; 41 T; 0 other;

Alignment Scores:

Pred. No.:	3.59e-07	Length:	243
Score:	170.00	Matches:	32
Percent Similarity:	65.82%	Conservative:	20
Best Local Similarity:	40.51%	Mismatches:	27
Query Match:	7.10%	Indels:	0

DB: 24 Gaps: 0
 US-09-912-020-325 (1-477) x ABN77217 (1-243)
 QY 378 AlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMet 397
 Db 1 GCGTCGGTGAAGCGCTGAAAGGGCCCGCCGCTGCAGGACGAGACGCCCGAGCC 60
 QY 398 IleValLeuGlyAlaLeuGluAlaValAspTrpValValSerPheGluGluAspThrPro 417
 Db 61 GCCGTATTGGCGAGCATTAAGGGCGTGGGATGTCAGCTTTTCGACGAAGATACGCCG 120
 QY 418 GlnArgLeuIleAlaGlyIleLeuProAspLeuLeuValLysGlyGlyAspTrpLysPro 437
 Db 121 ATCGAGCTCTCGCGCGCTCTCGCGGACGTGATCGTCAAGGGAGCGCACTATCGAGAA 180
 QY 438 GluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluValLeuValLeu 456
 Db 181 GACCAAGTGGTGGCGCGCGACCTTGTGAAGAAACACGAGCGAGGCTGCTACTTGTG 237
 RESULT 2
 ID AAT67445
 AC AAT67445 standard; DNA; 249 BP.
 AC AAT67445;
 DT 09-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein ORF 21742157.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 XX diagnosis; ds.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 CDS 1..249
 FT /*tag= a
 FT /note= "no stop codon given"
 XX
 PN WO9640893-A1.
 XX 19-DEC-1996.
 PD
 XX 06-JUN-1996; 96WO-US09122.
 PF
 XX 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 DR WPI; 1997-0523306/05.
 DR P-PSDB; AAW20221.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 XX infection, and to detect Helicobacter
 PS Claim 9; Page 176; 1481pp; English.
 CC
 CC This sequence encodes a H. pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

XX
SQ Sequence 249 BP; 84 A; 39 C; 62 G; 64 T; 0 other;

Alignment Scores: 2.03e-06 Length: 249
Pred. No.: 162.00 Matches: 35
Score: 65.00% Conservative: 17
Percent Similarity: 43.75% Mismatches: 24
Best Local Similarity: 6.77% Indels: 4
Query Match: 18 Gaps: 1

US-09-912-020-325 (1-477) x AAT67445 (1-249)

QY 394 GluGlnArgMetIleValLeuGlyAlaLeuGluAlaValAspTrpValValSerPheGlu 413
Db 10 AAAGACAGGCGGTTCTTTAGCGAGCTTCTTCGCGGATTAATGTTGGTGGGA 69
QY 414 GluAspThrProGlnArgLeuIleAlaGlyIleLeuProAspLeuValLysGlyGly 433
Db 70 GAAGACACGCCCATAAATTTGATTCAAGCCCTAAAGCCTGATATTTTAGTCAAGGGAGCG 129
QY 434 AspTyrLysProGluGluIleAlaGlySerLysGluValTrpAlaAsnGlyGlyGluVal 453
Db 130 GACTACCTCAATAAAGAGTCATAGGAGCGAG-----TTGGCTAAAGAAACC 177
QY 454 LeuValLeuAsnPheGluAspGlyCysSerThrThrAsnIleIleLysLysIleGlnGln 473
Db 178 CGTTGTAGAAATTTGAAGAGGTATTCCACAGCCCTATCATAGAAAAATTAAGG 237

RESULT 3

ID AAH81290/c
ID AAH81290 standard; DNA; 272 BP.

XX AAH81290;

XX 21-SEP-2001 (first entry)

XX Escherichia coli nucleotide sequence SEQ ID NO:89.

XX Escherichia coli; identification: proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition; ds.

XX Escherichia coli.

XX WO200148209-A2.

XX 05-JUL-2001.

XX 19-DEC-2000; 2000WO-US34419.

XX 23-DEC-1999; 99US-0173005.

XX (ELIT-) ELITRA PHARM INC.

XX Forsyth RA, Ohlsen KL, Zyskind JW;

XX WPI; 2001-457376/49.

XX Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents -

XX Claim 1; Page 128; 596pp; English.

XX The present invention describes a purified or isolated nucleic acid

CC sequence (1) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.

CC (1) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (1) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools;
CC for example, nucleic acid probes complementary to proliferation-required
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.

XX
SQ Sequence 272 BP; 65 A; 61 C; 61 G; 85 T; 0 other;

Alignment Scores: 0.017 Length: 272
Pred. No.: 120.00 Matches: 23
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 5.01% Indels: 0
Query Match: 22 Gaps: 0
DB:

US-09-912-020-325 (1-477) x AAH81290 (1-272).

QY 455 ValLeuAsnPheGluAspGlyCysSerThrThrAsnIleIleLysLysIleGlnGlnAsp 474
Db 271 GTGCTCAACTTTGAAGACGGTTGCTCGACGCCCAACATCATCAAGAAAGATCCACAGGAT 212

QY 475 LysLysGly 477

Db 211 AAAAAAGGC 203

RESULT 4

ABK63357/C

ID ABK63357 standard; cDNA; 292 BP.

XX ABK63357;

XX 18-JUN-2002 (first entry)

XX Rat sequence differentially expressed in response to a hepatotoxin #1264.

XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW differential expression; centrilobular necrosis; steatosis.

XX Rattus norvegicus.

XX WO200210453-A2.

XX 07-FEB-2002.

XX 30-JUL-2001; 2001WO-US23872.

XX 31-JUL-2000; 2000US-222040P.

XX 02-NOV-2000; 2000US-244880P.

XX 11-MAY-2001; 2001US-290029P.

XX 15-MAY-2001; 2001US-290645P.

XX 22-MAY-2001; 2001US-292336P.

XX 06-JUN-2001; 2001US-295798P.

XX 13-JUN-2001; 2001US-297457P.

XX 19-JUN-2001; 2001US-298884P.

XX 09-JUL-2001; 2001US-303459P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

XX WPI; 2002-241625/29.

CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA12532 to AA21779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX Sequence 300 BP; 70 A; 89 C; 92 G; 49 T; 0 other;

Alignment Scores:
 Pred. No.: 576 Length: 300
 Score: 71.50 Matches: 21
 Percent Similarity: 46.32% Conservative: 23
 Best Local Similarity: 22.11% Mismatches: 42
 Query Match: 2.99% Indels: 9
 DB: 20 Gaps: 3

US-09-912-020-325 (1-477) x AA214636 (1-300)

Qy 200 GluAlaValValGlyLysCysLysThrGluGluLeuValGluArgGlyMetLysLeu 219
 Db 21 GAGCAGACCCAGCGCTCTTAGCAATCCCGGCTGTGTGAAGCAGCGAAGCAGATG 80
 Qy 220 IleAlaAspTyrGluLeuSerAlaLeuValThrArgSerGluGluGlyMetSerLeu 239
 Db 81 GTCGGAGGCCAGCAACTACCTGCACCTT-----GCCGCCAAGAGTGGCAATCTTTT 131
 Qy 240 LeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspVal 259
 Db 132 AGGTCTCTCGGAGGCCGCCAGCTCCCTCCCACTGAGAAAGAGAGTTGGTA---ACC 188
 Qy 260 ThrGlyAlaGlyAspThrValIleGlyValLeuAlaAla-----ThrLeu 274
 Db 189 ACAGAGCAAGATCCCTGCGGCTGGGAAAGAGCTCATCACGGCGAGGCTCTGGGCCCAT 248
 Qy 275 AlaAlaGlyAsnSerLeuGluGluAlaCysPheAlaAsnAla 289
 Db 249 GTGGCTGGACAGACTGGGCACAGTCTTGTGGTCTGCTGGGAGC 293

RESULT 8

AAH81291/c

ID AAH81291 standard; DNA; 245 BP.

XX AC AAH81291;

XX 21-SEP-2001 (first entry)

XX Escherichia coli nucleotide sequence SEQ ID NO:90.

XX Escherichia coli; identification; proliferation; gene therapy; diagnosis;
 KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 KW bacterial growth inhibition; ds.

OS Escherichia coli.

XX WO200148209-A2.

XX 05-JUL-2001.

XX 19-DEC-2000; 2000WO-US4419.

XX 23-DEC-1999; 99US-0173005.

PA (ELIT-) ELITRA PHARM INC.

PI Forsyth RA, Ohlsen KL, Zyskind JW;

XX WPI; 2001-457376/49.

XX Novel nucleic acids encoding proteins required for Escherichia coli
 PT proliferation, useful for screening for antimicrobial agents -

XX Claim 1; Page 129; 596pp; English.

XX The present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
 CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.
 CC For example, nucleic acid probes complementary to proliferation-required
 CC sequences that are specific for particular species of microorganisms can
 CC be used as probes to identify particular microorganism species in
 CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
 CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
 CC represent oligonucleotides, which are used in the exemplification of the
 CC present invention.

XX Sequence 245 BP; 57 A; 53 C; 55 G; 80 T; 0 other;

Alignment Scores:
 Pred. No.: 610 Length: 245
 Score: 70.00 Matches: 14
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.93% Indels: 0
 DB: 22 Gaps: 0

US-09-912-020-325 (1-477) x AAH81291 (1-245)

Qy 464 ThrThrAsnIleIleLysLysIleGlnGlnAspLysLysGly 477
 Db 244 ACGACCAACATCATCAGAGAGATCCACAGGATAAAAAGGC 203

RESULT 9

ABL74436

ID ABL74436 standard; CDNA; 287 BP.

XX AC ABL74436;

XX 14-MAY-2002 (first entry)

XX Corn tassel-derived polynucleotide (cdps) SEQ ID NO:3810.

XX Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
 KW inheritance; characteristic; growth; development; disease resistance;
 KW environmental adaptability; quality; yield; molecular marker;
 KW multigene trait; plant breeding; corn tassel; gene; ss.

OS Zea mays.

XX US2001051335-A1.

XX 13-DEC-2001.

XX 16-APR-1999; 99US-0294093.

XX 21-APR-1998; 98US-082567P.

KW detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW gene; ds.
XX
OS Mycobacterium bovis.
XX
PN W0954487-A2.
XX
XX
PD 28-OCT-1999.
XX
XX 16-APR-1999; 99WO-1B00740.
XX
XX 16-APR-1998; 98US-0060756.
PR
XX (INSP) INST PASTEUR.
XX
XX Cole S, Buchrieser-Brosch R, Gordon S, Billault A;
PI WPI; 2000-013262/01.
XX
XX Isolation of polynucleotides from mycobacterial genomes, useful for
PT detection of Mycobacteria and for combating tuberculosis .
XX
PS Claim 27; Page 127; 161pp; English.
XX
CC The present invention describes a method for isolating a polynucleotide
CC of interest that is present or is expressed in a genome of a first
CC mycobacterium strain and that is absent or altered in a genome of a
CC second mycobacterium strain, which is different from the first strain
CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
CC vectors, which are preferably immobilised, can be used to detect
CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
CC samples. The polynucleotides identified are useful as probes or primers
CC for detecting a given mycobacterium of interest. By aligning the
CC polynucleotides contained in the recombinant BAC vectors it is possible
CC to physically map a polynucleotide of mycobacterial origin in a
CC biological sample. The methods and vectors from the present invention
CC are useful in providing information for combating tuberculosis. It is
CC possible to compare genomes between different strains or species and
CC their non-pathogenic strains or species counterparts. ABQ62492 to
CC ABQ63228 and AB81227 to AB81230 represent sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 279 BP; 37 A; 94 C; 105 G; 43 T; 0 other;

Alignment Scores:
Pred. No.: 1.88e+03 Length: 279
Score: 65.50 Matches: 22
Percent Similarity: 46.67% Conservative: 6
Best Local Similarity: 36.67% Mismatches: 27
Query Match: 2.74% Indels: 6
DB: 21 Gaps: 2

US-09-912-020-325 (1-477) x ABQ63223 (1-279)
QY 310 AsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGluLeuLys 329
Db 110 AATCTGGCGCTGGCGGACACCGCTGGCGGTAGGCTTG-----CGATCGTGCAGC 160
QY 330 LeuAlaValAlaAlaAlaArgLysArg-----GlyGluLysValValMetThrAsnGly 347
Db 161 GCTGGCGTGGCGGAGGAGATCCGACGAGATTGGGCG-AGATGCGGTGCTCACCATCGGG 219
QY 348 ValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaArgLysLeuGly 367
Db 220 GTATTTGACGGCTGGCACCGCGGCGACCGCACTGATCGCGCACGGGTCAAGGGCGC 279
RESULT 12
AAH52260
ID AAH52260 standard; cDNA; 237 BP.
XX
AC AAH52260;
XX
DT 10-SEP-2001 (first entry)

XX Human AFP protein encoding cDNA sequence SEQ ID NO:335.
DE
XX
KW Human; secreted protein; secretion; bacterial cell; fungal cell;
KW eukaryotic cell; fusion protein; maltose binding protein;
KW immunoglobulin constant region; polyhistidine tag; ss.
XX
OS Homo sapiens.
XX
XX WO200129221-A2.
XX
XX 26-APR-2001.
XX
XX 20-OCT-2000; 2000WO-US29052.
XX
XX 20-OCT-1999; 99US-0160712.
PR
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Conklin DC, Yee DP;
PI
XX
XX WPI; 2001-300340/31.
DR
DR P-PSDB; AAG81409.
XX
XX Isolated polypeptide for directing secretion of proteins of interest
XX from a host cell including, e.g. bacteria, includes contiguous amino
XX acid residues of polypeptide with specified amino acids .
PS Claim 9; Page 556; 617pp; English.
XX
XX AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
XX to AAG81453. The secreted proteins can be used for directing the
XX secretion of proteins of interest from a host cell including bacteria,
XX fungal cells, and cultured higher eukaryotic cells. The present invention
XX also describes fusion proteins, where a secreted protein of the invention
XX is operably linked via a peptide bond or peptide linker to a second
XX protein selected from the group consisting of maltose binding protein,
XX an immunoglobulin constant region, a polyhistidine tag and a peptide
XX given in AAG81453.
SQ Sequence 237 BP; 65 A; 41 C; 69 G; 62 T; 0 other;

Alignment Scores:
Pred. No.: 2.09e+03 Length: 237
Score: 64.00 Matches: 24
Percent Similarity: 46.75% Conservative: 12
Best Local Similarity: 31.17% Mismatches: 35
Query Match: 2.67% Indels: 6
DB: 22 Gaps: 3

US-09-912-020-325 (1-477) x AAH52260 (1-237)
QY 130 ProGlnProLeuHisGluArgIleAsnGlnAlaLeuSerIleGlyAlaLeuValLeu 149
Db 4 CCTGTGGTCTCTGCGCAAGAGGTGGAGTCGCTCTT-----GGGGTGGTGGTCTGTCTG 57
QY 150 SerAspTyrAlaLysGlyAlaLeuAlaSerValGlnGlnMetIleGlnLeuAlaArgLys 169
Db 58 GGTGGCTGTGCCTCAGGGGATTTTCGCTGTGTACAGGAAGCAATGGCAAAATGAGCAA 117
QY 170 AlaGly-----ValProValLeuIleAspProLysGlyThrAspPheGluArgTyr 186
Db 118 GTTGGGAAGTGTGTTCGCGAGACTACAGGATAAAAATACTATGAT---AAGAAATAC 174
QY 187 ArgGlyAlaThrLeuLeuThrProAsnLeuSerGluPheGluAlaValVal 203
Db 175 CAAGTATTCTGAAGCTGGTTGAACACCAAGAGGATTTTGGCGATCATG 225
RESULT 13
ABL38371
ID ABL38371 standard; cDNA; 275 BP.
XX
XX ABL38371;

```
XX 08-APR-2002 (first entry)
XX Human colon tumour antigen polynucleotide SEQ ID NO:1960.
XX DE
XX DE
XX Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
XX KW colon tumour metastatic antigen; diagnosis; gene; ss.
XX KW
XX OS Homo sapiens.
XX XX
XX W0200196388-A2.
XX PN
XX 20-DEC-2001.
XX PD
XX 08-JUN-2001; 2001WO-US18557.
XX PF
XX 09-JUN-2000; 2000US-210899P.
XX PR
XX 20-FEB-2001; 2001US-270216P.
XX PD
XX (CORI-) CORIXA CORP.
XX PA
XX Jiang Y, Harlocker SL, Secrist H;
XX PI WPI; 2002-114514/15.
XX DR
XX Novel isolated colon tumor polynucleotide differentially expressed in
XX PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
XX PT useful for inhibiting development of cancer in patient -
XX XX
XX Claim 1; SEQ ID 1960; 105pp; English.
XX PS
XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
XX CC which were isolated from human colon tumour and colon metastatic tumour
XX CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
XX CC production. (I) can be used for stimulating and/or expanding T cells
XX CC specific for a tumour protein on contact with the T cells. They are also
XX CC useful for inhibiting the development of cancer in a patient. (I) can be
XX CC used as probes or primers for nucleic acid hybridisation, for preparing
XX CC mutant species primers, or primers for use in genetic constructions. (I)
XX CC can be used in the diagnosis of a colon tumour.
XX XX
XX Sequence 275 BP; 97 A; 58 C; 64 G; 54 T; 2 other;
XX SQ
Alignment Scores:
Pred. No.: 2.53e+03 Length: 275
Score: 64.00 Matches: 23
Percent Similarity: 39.36% Conservative: 14
Best Local Similarity: 24.47% Mismatches: 29
Query Match: 2.67% Indels: 28
DB: 24 Gaps: 3
US-09-912-020-325 (1-477) x ABL38371 (1-275)
Qy 291 AlaGlyValValGlyLysLeuGlyThrSerThrValSerProileGluLeuGluAsn 310
Db 46 GCAGAAGTGGTCCAGGACAAAGTAGGCCCCAGCAATAAGTCCC----- 90
Qy 311 AlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGluLeuLysLeu 330
Db 91 -----CCCATTAAGAGAGAGAAACAAA--- 114
Qy 331 AlaValAlaAlaAlaArgLysArgGlyGluLysValValMetThrAsnGlyValPheAsp 350
Db 115 -----GGAGATTCGTAGAAAAAATCAAGATTACTATGAC 150
Qy 351 -----lleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaArgLysLeuGly 367
Db 151 ATGAATCCATGGTCCATGCAGACACAAAGATCATTTATTCTGAAGAGCCAAAGCTGCT 210
Qy 368 AspArgLeuIleValAlaValAsnSerAspAlaSerThrLys 381
Db 211 GAGGAGTAGTAGTGGCCACCACCAACCAAGAGTCGGGGATGAAG 252
```

```
RESULT 14
ABL76194/C
ID ABL76194 standard; cDNA; 298 BP.
XX
XX ABL76194;
AC
XX
XX 14-MAY-2002 (first entry)
XX DT
XX Corn tassel-derived polynucleotide (cdps) SEQ ID NO:5568.
XX DE
XX Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPS;
XX KW inheritance; characteristic; growth; development; disease resistance;
XX KW environmental adaptability; quality; yield; molecular marker;
XX KW multigene trait; plant breeding; corn tassel; gene; ss.
XX XX
XX Zea mays.
XX OS
XX US2001051335-A1.
XX PN
XX 13-DEC-2001.
XX PD
XX 16-APR-1999; 99US-0294093.
XX PF
XX 21-APR-1998; 98US-082567P.
XX PR
XX (LALG/) LALGUDI R V.
XX PA (ITOL/) ITO L Y.
XX PA (SHER/) SHERMAN B K.
XX XX
XX Lalgudi RV, Ito LY, Sherman BK;
XX PI WPI; 2002-163647/21.
XX DR
XX Novel purified corn tassel-derived polynucleotide useful for
XX PT determining altered gene expression, to recover regulatory elements and
XX PT to follow inheritance of desirable characteristics through hybrid
XX PT breeding programs -
XX XX
XX Claim 1; SEQ ID 5568; 201pp; English.
XX PS
XX The present sequence describes a purified corn tassel-derived
XX CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
XX CC selected from those given in ABL76194 to ABL76833. The cdps sequences
XX CC encode corn tassel-derived polypeptides (CDPS). The cdps sequences (I)
XX CC can be used for determining altered gene expression, to recover
XX CC regulatory elements and to follow inheritance of desirable
XX CC characteristics through hybrid breeding programs. (I) are also useful
XX CC in the evaluation, and alteration of desired characteristics associated
XX CC with growth and development, disease resistance, environmental
XX CC adaptability, quality and yield, and as molecular markers for studying
XX CC inheritance of multigene traits in a plant breeding program. (I) can be
XX CC used to produce a tassel-specific profile of gene transcription, a
XX CC transcript image, to clone regulatory elements for use in transformation
XX CC vectors, to express a polypeptide, to identify, isolate or extend
XX CC identical or related corn tassel nucleic acid sequences from DNA
XX CC libraries, in nucleic acid hybridisation or amplification technologies,
XX CC as query sequences to determine homology of known sequences, as probe
XX CC for use in Southern or Northern hybridisation, and to identify the
XX CC presence of and/or to determine the degree of similarity between two
XX CC (or more) nucleic acid sequences.
XX SQ Sequence 298 BP; 55 A; 87 C; 87 G; 65 T; 4 other;
```

```
Alignment Scores:
Pred. No.: 2.81e+03 Length: 298
Score: 64.00 Matches: 25
Percent Similarity: 37.50% Conservative: 14
Best Local Similarity: 24.04% Mismatches: 47
Query Match: 2.67% Indels: 18
DB: 24 Gaps: 3
```

US-09-912-020-325 (1-477) x ABL76194 (1-298)

```
QY 28 ProThrSerArgile-----SerProGluAlaProValProValValValValValAsnThr 45
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 CCAATCAGCAGCGCCCTCGGGGAACCCAGGCAATCTCCTATTAGCACCTCAATGCT 238
QY 46 IleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSerLeuGlyAla 65
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 GTNAATTCATCAATGACATTCCTGCCAATGTCGACAGCAAAAGTCGCCACGGATGGTGC 178
QY 66 AsnAlaArgLeuValGlyLeuThrGlyIleAspAspAlaAlaArgAlaLeuSerLysSer 85
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 GGGCTCAGAACCGAGGGGTTGGTGGCCCAATGATCTTGGCCGCGCATGTGTGACACGCGT 118
QY 86 LeuAlaAspValAsnValLysCysAspPheValSerValProThrHisProThrIleThr 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 117 CTTGCC-----CTCCAGACCATGGCCAC----- 94
QY 106 LysLeuArgValLeuSerArgAsnGlnGlnLeuIleArgLeuAspPheGluGluGlyPhe 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93 -----CACAGGCCAGAGATGATGTAGTCCACGAGGCCCTGGAGAGAGGGCTT 46
QY 126 GluGlyValAsp 129
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 GGAGGGCGCAGAT 34
RESULT 15
ID AAK17810/c
XX AAK17810;
AC AAK17810;
XX
XX
XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe SEQ ID NO: 17801.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains .
XX
XX Example 4; SEQ ID NO: 17801; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
```

```
SQ Sequence 267 BP; 66 A; 82 C; 55 G; 64 T; 0 other;
Alignment Scores: 2.71e+03 Length: 267
Pred. No.: 63.50 Matches: 23
Score: 44.44% Conservative: 13
Percent Similarity: 28.40% Mismatches: 28
Best Local Similarity: 2.65% Indels: 17
Query Match: 22 Gaps: 3
DB:
US-09-912-020-325 (1-477) x AAK17810 (1-267)
QY 23 ArgTyrTrpTyrGlyProThrSerArgIleSerProGluAlaProValProValValLys 42
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 260 AAGGTATGGCACCACCAATGTACCCAGAGGATATCTGGCAGGTGCACAGTACACATAATAGTC 201
QY 43 ValAsnThrIleGluGluArgProGlyGlyAlaAlaAsnValAlaMetAsnIleAlaSer 62
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 CACATAAATAGTAGCAACCCAGA-----GCCACCGTAGCCATCTGGCAGAGCAGAC 153
QY 63 LeuGlyAlaAsnAla-----ArgLeuValGlyLeu 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 TTGGAGTGCAGCTCTGTGAGTCACAGTTGGGGCTTCCTGTGTAACTGGTGAGGCTC 93
QY 73 ThrGlyIleAsp-----AspAlaAlaArgAlaLeuSerLysSerLeuAlaAspVal 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 ACAGGAGTCCAATGGCTTAGAGTGGCAGATAGGGCCATGAGTCGGGCCATTGTCTCTGTG 33
QY 90 Asn 90
   |||
Db 32 TCA 30
```

Search completed: November 24, 2002, 23:26:15
Job time : 296 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 24, 2002, 23:19:40 : Search time 2366 Seconds
(without alignments)
3265.109 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERAGVMVGVDM.....PEDGCGTNIKKIQDKKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 6209604

Minimum DB seq length: 0
Maximum DB seq length: 300

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-Q/cgn2_l/USPTO.spool/US09912020/runat_20112002_150300_11098/app_query.fasta_1.647
-DB-EST -QFMT-fastap -SUFFIX-p2nszlm300.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=biosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=300
-USER=US09912020 @CGN_1_1_2024 @runat_20112002_150300_11098 -NGPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_nam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	123	5.1	263	9	AU179292	AU179292 AU179292
2	116	4.8	182	9	AA824092	AA824092 VF66404.S
3	112	4.7	292	9	AU1236089	AU1236089 EST232651
4	110.5	4.6	240	9	AU072448	AU072448 AU072448
5	108	4.5	238	14	H32797	H32797 EST108244.R
6	104.5	4.4	240	9	AU074294	AU074294 AU074294
7	98	4.1	300	9	AU100188	AU100188 AU100188
8	97.5	4.1	235	14	BM903102	BM903102 NXLV_079
9	97	4.1	242	12	BF755713	BF755713 PM4-CT056
10	93	3.9	282	10	BB156904	BB156904 BB156904
11	92	3.8	150	10	AW565187	AW565187 LGL_328.B
12	88	3.7	297	10	AW782015	AW782015 sl99e02.Y
13	84	3.5	249	14	BQ634379	BQ634379 NXR068.C
14	83.5	3.5	268	14	BQ697625	BQ697625 NXPV_058
15	82	3.4	260	9	AA113615	AA113615 0662R.PYR
16	81.5	3.4	289	14	BQ041062	BQ041062 qd33c07.Y
17	80	3.3	256	13	BI200260	BI200260 nlb09fs.r
18	79	3.3	241	13	BM158266	BM158266 NXLV_031
19	79	3.3	287	13	BI129263	BI129263 G088P18Y
20	78.5	3.3	270	10	BB355746	BB355746 BB355746
21	78	3.3	236	14	W17420	W17420 mb58h11.r1
22	78	3.3	288	14	BQ635002	BQ635002 NXR075.G
23	76.5	3.2	286	12	BF659338	BF659338 maal4c03
24	76.5	3.2	295	10	AV413900	AV413900 AV413900
25	75.5	3.2	224	12	CG157882	CG157882 EML1_6-A10
26	75.5	3.2	300	14	C55835	C55835 C55835 YUJ1
27	74.5	3.1	271	14	C57175	C57175 C57175 YUJ1
28	74	3.1	247	12	BF226625	BF226625 uz05d07.X
29	73.5	3.1	246	9	AV083373	AV083373 RST10248
30	73.5	3.1	251	12	BG191033	BG191033 EST10248
31	73.5	3.1	266	9	AA432985	AA432985 V687C09.S
32	73.5	3.1	269	9	AI974677	AI974677 T11332e
33	73	3.1	110	9	AI616187	AI616187 VF66404.X
34	73	3.1	243	12	BF849567	BF849567 IL5-EN008
35	72	3.0	290	12	BF564215	BF564215 UI-R-C4-a
36	71.5	3.0	279	13	BI423558	BI423558 949049F03
37	71.5	3.0	292	12	BE854380	BE854380 UX29F02.Y
38	71	3.0	226	9	AA356818	AA356818 EST65690
39	71	3.0	285	9	AI165611	AI165611 A087P07u
40	71	3.0	298	9	AI426060	AI426060 mh29d04.X
41	70.5	2.9	218	12	BF585739	BF585739 FM1_24.A0
42	70.5	2.9	218	13	BM418225	BM418225 952004F03
43	70.5	2.9	228	9	AI163266	AI163266 A038d4u
44	70.5	2.9	256	10	AV552483	AV552483 AV552483
45	70.5	2.9	258	14	BQ046251	BQ046251 EST595369

ALIGNMENTS

RESULT 1
AU179292
LOCUS
DEFINITION
AU179292 Medaka liver cDNA library (Ole) from HNI Oryzias latipes
CDNA clone Ole06.08a similar to pir173720| ethanolamine-phosphate
cytidyltransferase (EC 2.7.7.14) - fission yeast
(Schizosaccharomyces pombe), mRNA sequence.
ACCESSION
AU179292
VERSION
AU179292.1 GI:13428129
KEYWORDS
EST.
SOURCE
Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;

Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source Location/Qualifiers

1..292
/organism="Rattus sp."
/db_xref="ATCC (inhost):2041486"
/db_xref="taxon:10118"
/clone="ROVDB29"
/clone_lib="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 67 a 60 c 86 g 79 t

ALIGNMENT SCORES:

Pred. No.: 0.0062 Length: 292
Score: 112.00 Matches: 29
Percent Similarity: 54.93% Conservative: 10
Best Local Similarity: 40.85% Mismatches: 30
Query Match: 4.68% Indels: 2
DB: 9 Gaps: 1

US-09-912-020-325 (1-477) x AI236089 (1-292)

Qy 237 MetSerLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluVal 256
Db 278 GTGACACTGTACAGCGCAGACCTGTTCCAAAGCACATCCACAGAGCAGCTCAAGGCT 219
Qy 257 TyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAla--- 275
Db 218 GTGGACACACGGGTGCTGTGACAGTTTGTGGAGCGCTTGCTTACCTGCTTAC 159
Qy 276 ---AlaGlyAsnSerLeuGluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValVal 294
Db 158 TACCAAGTCTGCTCTTGAAGAAATGCTCAAGAGATCTAATTCATCGCTGCGTCAAGC 99
Qy 295 ValGlyLysLeuGlyThrSerThrValSerPro 305
Db 98 GTCCAGGCCACAGGACACAGCTCCTTTATCCA 56

RESULT 4
AU072448 240 bp mRNA linear EST 23-JUN-1999
LOCUS AU072448 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
DEFINITION discoideum cDNA clone SSE551, mRNA sequence.

ACCESSION AU072448
VERSION AU072448.1 GI:5162637

KEYWORDS EST.
SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE 1 (bases 1 to 240)

AUTHORS Urushihara, H.

TITLE Developmental cDNA in Dictyostelium discoideum (1999)

JOURNAL Unpublished (1999)

COMMENT Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Tel: 81-298-53-4664

Fax: 81-298-53-6614

Email: hideko@biol.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES Location/Qualifiers

1..240
/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SSE551"

/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev_stage="slug"

BASE COUNT 93 a 27 c 53 g 67 t

ORIGIN

Alignment Scores:
Pred. No.: 0.00672 Length: 240
Score: 110.50 Matches: 27
Percent Similarity: 52.00% Conservative: 12
Best Local Similarity: 36.00% Mismatches: 31
Query Match: 4.62% Indels: 5
DB: 9 Gaps: 2

US-09-912-020-325 (1-477) x AU072448 (1-240)

Qy 343 ValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsn 362
Db 24 GTTTATGTTGATGGTGTGTTTGAATTAATGCAATTTGGACATGCAATTAAGACAA 83
Qy 363 AlaArgLysLeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThrLysArg 382
Db 84 GCAGAGAAATTAGGAGATATTTTAGTAGTGTGTACATACAGATGAAGAAATTCGAAAA 143
Qy 383 LeuLysGlyAspSerArgProValAsnProLeuGluGlnArgMetIleValLeuGlyAla 402
Db 144 AATAAAGGTCCA-----CCAGTTATGATGAACAGAAAGA-----TATAAGCA 188

Qy 403 LeuGluAlaValAspTrpValValSerPheGluGluAspThrPro 417

Db 189 GTACGTGCATGTAATGGCAGATGAAGTTGCAGAAAGGTGCACCT 233

RESULT 5

H32797

LOCUS H32797

DEFINITION H32797

EST. Rattus sp.

ACCESSION H32797

VERSION H32797.1 GI:978214

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 238)

AUTHORS Lee, N.H., Weinstock, K.G., Kirkness, E.F., Earle-Hughes, J.A., Fuldner,

R.A., Marmaras, S., Glodek, A., Gocayne, J.D., Adams, M.D., Kerlavage

, A.R., Fraser, C.M. and Venter, J.C.

Comparative expressed-sequence-tag analysis of differential gene

expression profiles in PC-12 cells before and after nerve growth

factor treatment

Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)

95396786

CONTACT: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

For clone availability please contact the TIGR Database

(tdbinfo@tdb.tigr.org)

Seq primer: M13 Reverse.

FEATURES Location/Qualifiers

1..238
/organism="Rattus sp."

/db_xref="ATCC (inhost):200292"

/db_xref="taxon:10118"

/clone="RPCJ75"

/clone_lib="Rat PC-12 cells, untreated"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; poly(A)+ RNA was purified from untreated PC12 cells

cultured for 9 days. cDNA was constructed using an

oligo-dT primer and directionally cloned using the Lambda

ZAP II Vector kit by Stratagene"

BASE COUNT 55 a 63 c 68 g 52 t

```
ORIGIN
Alignment Scores:
Pred. No.: 0.013 Length: 238
Score: 108.00 Matches: 25
Percent Similarity: 50.00% Conservative: 12
Best Local Similarity: 33.78% Mismatches: 35
Query Match: 4.51% Indels: 2
DB: 14 Gaps: 1

US-09-912-020-325 (1-477) x H32797 (1-238)
QY 339 GlyLysValValMetThrAsnGlyValPheAspLeuHisAlaGlyHisValSer 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GGGAGACAGTCATCTATGTGGCTGTGCTGCTTTAACTGTTCCACATCGGCGACGTGGAC 60

QY 359 TyrLeuAlaAsnAlaArgLysLeuGlyAspArg-----LeuLeValaValaValaSer 376
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TTCTCAGAGAGGTGCACAGCTAGCCAGAGCCCTACGTCATCGTGGCGCTACACATT 120

QY 377 AspAlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGluGlnArg 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GACCAGGAAGTAACCGGTACAGGCGAGAACTACCCCATGTAACCTGCATGACGCG 180

QY 397 MetLeuValLeuGlyAlaLeuGluAlaValaValaAspTrpValVal 410
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 ACTCTCAGTGTGCTGCGCTGCGGTATGTTTCAGAAAGTGCTG 222

RESULT 6
AU074294
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 0.0336 Length: 240
Score: 104.50 Matches: 25
Percent Similarity: 56.06% Conservative: 12
Best Local Similarity: 37.88% Mismatches: 24
Query Match: 4.37% Indels: 5
DB: 9 Gaps: 2

US-09-912-020-325 (1-477) x AU074294 (1-240)
QY 343 ValMetThrAsnGlyValPheAspLeuHisAlaGlyHisValSerTyrLeuAlaAsn 362

US-09-912-020-325 (1-477) x AU100188 (1-300)
QY 311 AlaValArgGlyArgAla-----AspThrGlyPheGlyValMetThrGlu 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 GCTGTCCGCCGCCGCGTGTGCGGGGCCATGATCGGAACGCGGCGTCTGAGCGGCG 85

QY 326 GluGluLeuLysLeuAlaValaValaAlaAlaArgLysArgGlyGluLysValMetThr 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 CAGAGCAGCGCGCGCGGCGGCGGCGCGCC-GTGAGG-----GTGTGTGTC 132
```

```

QY 346 AsnGlyValPheAspIleLeuHisAlaGlyHisValSerTyrLeuAlaAsnAlaAArgLys 365
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 133 GATGGCTGCTATGACATGTGTACGACCCATCCACACCGAGTGGCGGAGGCGGCC 192
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 366 LeuGlyAspArgLeuIleValAlaValAsnSerAspAlaSerThrLysArgLeuLysGly 385
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 193 ATGGGTGACATGCTGCTGCGTGCACCGCATGAGGAGATGCCCAAGCACACAGGGG 252
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 386 AspSerArgProValAsnProLeuGluGlnArgMetIleValLeuGlyAlaLeu 403
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 253 CCC-----CCGGTGTCTACTCAGGAGGAGATACACAGATGGTCAGGCCATC 300
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

RESULT 8
LOCUS      BM903102
DEFINITION NLV_079_E07_F_NXLV (Nsf Xylem Late wood vertical) Pinus taeda cDNA
ACCESSION  BM903102
VERSION     BM903102.1 GI:19385890
KEYWORDS   EST.
SOURCE     loblolly pine.
ORGANISM   Pinus taeda.
REFERENCE  1 (bases 1 to 235)
AUTHORS    Sederoff,R.
TITLE      Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL    Unpublished (2000)
COMMENT    Contact: Johnson, Arthur
           North Carolina State University
           Tel: 919 515 7800
           Fax: 919 515 7801
           Email: ajohnson@unity.ncsu.edu
           Seq primer: T3.

FEATURES             Location/Qualifiers
     source           1..235
                     /organism="Pinus taeda"
                     /strain="Coastal plain loblolly pine from North Carolina"
                     /db_xref="taxon:3352"
                     /clone="NXLV_079_E07"
                     /clone_lib="NXLV_079_E07"
                     /tissue_type="primary xylem"
                     /dev_stage="late wood"
                     /lab_host="XLL-Blue"
                     /note="Vector: pTriplex; Site 1: EcoRI. The library is
                     from late (summer-August) wood, taken from below the crown
                     of a 20 year old tree. The harvested xylem tissue was on
                     the cusp between transitional and mature wood. NOTE:
                     The sequences contain a 'cDNA adapter' between the EcoRI
                     site and the start of the EST. The adapter sequence is
                     'AATTCGCCCATTTATGGCC'."

BASE COUNT    65 a 36 c 55 g 71 t 18 others
ORIGIN
Alignment Scores:
Pred. No.:      0.213      Length:      235
Score:          97.50      Matches:      22
Percent Similarity: 48.44%      Conservative: 9
Best Local Similarity: 34.38%      Mismatches: 24
Query Match:    14         Indels:      9
DB:             1         Gaps:         1

US-09-912-020-325 (1-477) x BM903102 (1-235)

QY 236 GlyMetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGlu 255
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3  GCAATTCGGCCATTATTCGGCGGGGAG-----GAAAT 35
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 256 ValTyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAla 275
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 36  ATGTGTCATACAAATGCACGACGATGCTGTTGTAGGAGGCTTTTACTCAATGGTGG 95
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 276 AlaGlyAsnSerLeuGluAlaLacysPhePheAlaAsnAlaAlaGlyValVal 295

```

```

Db 96  CTAGGTAAATCTATTGAGGAATGTGTCAAAATATGGAATATTATCGGCAAAATGTTATCAT 155
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 296 GlyLysLeuGly 299
      ::::: :::::
Db 156 CAACGATCTGGT 167
      :::::

RESULT 9
LOCUS      BF755713/c
DEFINITION PM4-CT0562-291000-002-g10 CT0562 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF755713
VERSION     BF755713.1 GI:12103613
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 242)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
           Nagai,M.A., de Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
           Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
           Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
           ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
           Simpson,A.J.
           Shotgun sequencing of the human transcriptome with ORF expressed
           sequence tags
           Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
           20202663
           Contact: Simpson A.J.G.
           Laboratory of Cancer Genetics
           Ludwig Institute for Cancer Research
           Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
           Brazil
           Tel: +55-11-2704922
           Fax: +55-11-2707001
           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPESP/LICR Human Cancer Genome
           project. This entry can be seen in the following URL
           (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-CT0562-
           291000-002-g10&t3=2000-10-29&t4=1)
           Seq primer: puc 18 forward
           High quality sequence start: 11
           High quality sequence stop: 242.

FEATURES             Location/Qualifiers
     source           1..242
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0562"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."

BASE COUNT    50 a 70 c 63 g 59 t
ORIGIN
Alignment Scores:
Pred. No.:      0.255      Length:      242
Score:          97.00      Matches:      20
Percent Similarity: 50.68%      Conservative: 17
Best Local Similarity: 27.40%      Mismatches: 34
Query Match:    12         Indels:      2
DB:             1         Gaps:         1

US-09-912-020-325 (1-477) x BF755713 (1-242)

QY 324 ThrGluGluLeuLysLeuAlaValAlaAlaAArgLysArgGlyLysValVal 343
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 223 TCTCAGAAGATCATCCAGTTTCTTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 164
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

Qy	344	MetThrasnGlyValPheaspIleLeuHisAlaGlyHisValSerTyLeuAlaAsnAla	363
Db	163	TATGTGGCTGGCTCCACCTGTTCACATCGGCATGTGCATCTCTCGAAGAGTG	104
Qy	364	ArgLysLeuGlyAspArg-----LeuIleValAlaValAsnSerAspAlaSerThrLys	381
Db	103	CACAGGCTGGCAGAGAGGCCCTACATCTTCGCGGCTTGTGACCGAGGAGGTCAAT	44
Qy	382	ArgLeuLysGlyAspSerArgProValAsnProLeuGlu	394
Db	43	CACACAGGGGAGAGACTACCCCATCATCATCTGCAC	5
RESULT 10	BB156904	282 bp mRNA linear EST 29-JUN-2000	
LOCUS	BB156904	RIKEN full-length enriched, 16 days neonate thymus Mus	
DEFINITION	BB156904	musculus cDNA clone Al30033M23 3', mRNA sequence.	
ACCESSION	BB156904		
VERSION	BB156904.1	GI:8812834	
KEYWORDS		EST.	
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
		1 (bases 1 to 282)	
		P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,	
		Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,	
		Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,	
		Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.	
		Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.	
		Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata	
		Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.	
		Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomingawa, N., Toya	
		T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.	
		Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino	
		M., Muramatsu, M., and Hayashizaki, Y.	
		RIKEN Mouse ESTs (Konno, H., et al.)	
		Unpublished (2000)	
TITLE		Contact: Yoshihide Hayashizaki	
JOURNAL		Laboratory for Genome Exploration Research Group, RIKEN Genomic	
COMMENT		Sciences Center (GSC), Yokohama Institute	
		The Institute of Physical and Chemical Research (RIKEN)	
		1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan	
		Tel: 81-45-503-9222	
		Fax: 81-45-503-9216	
		Email: genome-res@gsc.riken.go.jp,	
		URL: http://genome.gsc.riken.go.jp/	
		Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki	
		N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
		Thermotranscription and thermoactivation of thermolabile enzymes by	
		trehalose and its application for the synthesis of full length	
		cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	
		Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,	
		Tomaru, Y., Carninci, P., Akiyama, J., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki	
		Y., and Hayashizaki, Y.	
		Automated filtration-based high-throughput plasmid preparation	
		system. Genome Res. 9 (5), 463-470 (1999)	
		Carninci, P. and Hayashizaki, Y.	
		High-efficiency full-length cDNA cloning. Methods Enzymol. 303,	
		19-44 (1999)	
		Please visit our web site (http://genome.rtc.riken.go.jp) for	
		further details.	
FEATURES		Location/Qualifiers	
source		1..282	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="Al30033M23"	
		/clone.lib="RIKEN full-length enriched, 16 days neonate	
		thymus"	
		/tissue_type="thymus"	
		/dev_stage="16 days neonate"	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 249)
 AUTHORS Sedotoff, R.
 TITLE Molecular Basis of Wood Formation in the Pine Megagenome
 JOURNAL Unpublished (2000)
 COMMENT Contact: Johnson, Arthur
 North Carolina State University
 Tel: 919 515 7800
 Fax: 919 515 7801
 Email: ajohnson@unity.ncsu.edu
 Seq primer: T3.

FEATURES
 source
 1. .249
 Location/Qualifiers
 /organism="Pinus taeda"
 /strain="Coastal plain loblolly pine from North Carolina"
 /db_xref="taxon:3352"
 /clone="NXRV068_C02"
 /clone_lib="NXRV (Nsf Xylem Root wood Vertical)"
 /tissue_type="Xylem"
 /cell_type="Root (primary)"
 /dev_stage="transitional"
 /lab_host="X11-Blue"
 /note="vector: Bluescript SK-; Site_1: Eco RI; Site_2: XhoI; The library is from primary xylem scraped from the roots of a twelve year old tree in the transitional phase from juvenile wood to mature wood production. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'AATTCGGCACGAG'."
 'AATTCGGCACGAG'."

BASE COUNT 48 a 54 c 72 g 70 t 5 others
 ORIGIN

Alignment Scores:
 Pred. No.: 8.77 Length: 249
 Score: 84.00 Matches: 27
 Percent Similarity: 46.07% Conservative: 14
 Best Local Similarity: 30.34% Mismatches: 26
 Query Match: 3.51% Indels: 22
 DB: 14 Gaps: 4

US-09-912-020-325 (1-477) x BQ634379 (1-249)

Qy 13 ValMetValValGlyAspValMetLeuAspArgTyrTrpTyrGlyProThrSerArgile 32
 Db 25 ATCGTTGCTTGGGAGATCTCATTCATTT-----
 Qy 33 SerProGluAlaProValProValValValysValAsnThrIleGluGluArgProGly--- 51
 Db 58 -----GTCCCCACGGTCTCGGATCTTTCTGGTGAAGCGCCGGATTC 102
 Qy 52 -----GlyAla---AlaAsnValAlaMetAsnIleAlaSerLeuGlyAla 65
 Db 103 CAAGAAGTCGCAAGTGGGACCTGCTAATGTGGCTTGGAAATTCAGCGTGGG 162
 Qy 66 AsnAlaArgLeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSer 85
 Db 163 CGATCCGATTTGTTGGCAAGTTGGGATGATGACCTTGTGCGATGCTCGCGTACATA 222
 Qy 86 LeuAlaAspValAsnValLysCysAsp 94
 Db 223 CTGACGGAA---AACATTGGATCGAC 246

RESULT 14
 BQ697625
 LOCUS BQ697625 268 bp mRNA linear EST 15-JUL-2002
 DEFINITION NXPV_058_H03_F NXPV (Nsf Xylem Planings wood Vertical) Pinus taeda
 cDNA clone NXPV_058_H03 5', mRNA sequence.
 ACCESSION BQ697625
 VERSION BQ697625.1 GI:21822941
 KEYWORDS EST.
 SOURCE loblolly pine.
 ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 268)
 AUTHORS Sedotoff, R.
 TITLE Molecular Basis of Wood Formation in the Pine Megagenome
 JOURNAL Unpublished (2000)
 COMMENT Contact: Johnson, Arthur
 North Carolina State University
 Tel: 919 515 7800
 Fax: 919 515 7801
 Email: ajohnson@unity.ncsu.edu
 Seq primer: T3.

FEATURES
 source
 1. .268
 Location/Qualifiers
 /organism="Pinus taeda"
 /strain="Coastal plain loblolly pine from North Carolina"
 /db_xref="taxon:3352"
 /clone="NXPV_058_H03"
 /clone_lib="NXPV (Nsf Xylem Planings wood Vertical)"
 /tissue_type="Xylem"
 /cell_type="Planings (secondary)"
 /dev_stage="transitional"
 /lab_host="X11-Blue"
 /note="vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI; The library is from early (spring) secondary wood, taken from a ten year old tree in the transitional phase. The tree is a kind gift of the Westvaco Corporation. Secondary xylem was harvested from the tree by peeling back the bark and primary xylem and then removing the underlying tissue with a block plane. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'AATTCGGCACGAG'."
 'AATTCGGCACGAG'."

BASE COUNT 61 a 45 c 78 g 76 t 8 others
 ORIGIN

Alignment Scores:
 Pred. No.: 11.3 Length: 268
 Score: 83.50 Matches: 21
 Percent Similarity: 55.36% Conservative: 10
 Best Local Similarity: 37.50% Mismatches: 24
 Query Match: 3.49% Indels: 1
 DB: 14 Gaps: 1

US-09-912-020-325 (1-477) x BQ697625 (1-268)

Qy 38 ValProValValLysValAsnThrIleGluGluArgProGlyGlyAla---AlaAsnVal 56
 Db 11 GTGTCATTGGCTGATGCTCTGTCATTCAGAAAGCTCCAGGGGAGCCCTCGCAATGTT 70
 Qy 57 AlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArgLeuValGlyLeuThrGlyIleAsp 76
 Db 71 GCCGTGTTATAGCAGGCTCGAGGTTCATCCGCATTTATAGGGAAGNTTGGCGAGGAT 130
 Qy 77 AspAlaAlaArgAlaLeuSerLysSerLeuAlaAspValAsnValLys 92
 Db 131 GAATTTGGCGCATGCTTGTGGATATCTGAAGAAAATAATGTNGAG 178

RESULT 15
 AAL13615/c
 LOCUS AAL13615 260 bp mRNA linear EST 08-NOV-1996
 DEFINITION 0662R Pyrococcus furiosus 1-2AP II library, F Robb Pyrococcus
 furiosus cDNA similar to TRANSFERASE, GLYCEROL-3-PHOSPHATE
 CVD1DLYL, mRNA sequence.
 ACCESSION AAL13615
 VERSION AAL13615.1 GI:1665570
 KEYWORDS EST.
 SOURCE Pyrococcus furiosus.
 ORGANISM Pyrococcus furiosus.
 Archaea: Euryarchaeota; Thermococci; Thermococcales;
 Thermococcaceae; Pyrococcus.
 REFERENCE 1 (bases 1 to 260)
 AUTHORS Borges, K.M., Brummet, S.R., Bogert, A., Davis, M.C., Hujer, K.M., Domke, S.T., Szasz, J., Ravel, J., DiRuggiero, J., Fuller, C., Chase, J.W. and

Search completed: November 25, 2002, 00:59:48
Job time : 2371 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 24, 2002, 23:20:29 ; Search time 74 seconds
(without alignments)
1976.822 Million cell updates/sec

Title: US-09-912-020-325

Perfect score: 2393

Sequence: 1 MKVTLPFERAGVMVGVDM.....FEDGCGTNIKKIQDKKG 477

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 727474

Minimum DB seq length: 0

Maximum DB seq length: 300

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09912020/runat_20112002_150300_11121/app_query.fasta_1.647
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2nszlm300.rni -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosu62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=300 -USER=US09912020.ecgn_1_1_38_@runat_20112002_150300_11121 -NCFU=6
-ICFU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	11.0	294	4	US-08-651-1558-186
2	65.5	2.7	280	4	US-09-060-756-722
3	61	2.5	294	2	US-08-716-942-22
4	61	2.5	294	2	US-09-130-337A-22
5	59.5	2.5	268	1	US-08-105-168B-4
6	59.5	2.5	268	2	US-08-698-948-4
7	59	2.5	236	4	US-09-060-756-209
8	58.5	2.4	290	4	US-08-818-112-37
9	58.5	2.4	290	4	US-08-818-111-37
10	58.5	2.4	290	4	US-09-056-556-37
11	58.5	2.4	290	4	US-09-072-596-37
12	57.5	2.4	155	4	US-08-818-112-39

c 13	57.5	2.4	155	4	US-08-818-111-39	Sequence 39, Appl
c 14	57.5	2.4	155	4	US-09-056-556-39	Sequence 39, Appl
c 15	57.5	2.4	155	4	US-09-072-596-39	Sequence 39, Appl
c 16	57.5	2.4	155	4	US-09-275-848-9	Sequence 9, Appl
c 17	57	2.4	239	4	US-09-115-407-52	Sequence 52, Appl
c 18	57	2.4	262	4	US-08-944-483-5	Sequence 5, Appl
c 19	56.5	2.4	276	1	US-08-554-161-1	Sequence 1, Appl
c 20	56	2.3	274	4	US-09-275-848-10	Sequence 10, Appl
c 21	54.5	2.3	268	1	US-08-105-168B-7	Sequence 7, Appl
c 22	54.5	2.3	268	2	US-08-698-948-7	Sequence 7, Appl
c 23	54.5	2.3	279	4	US-09-086-483A-8	Sequence 8, Appl
c 24	54	2.3	124	4	US-08-857-046A-20	Sequence 20, Appl
c 25	53.5	2.2	268	1	US-08-105-168B-5	Sequence 5, Appl
c 26	53.5	2.2	268	2	US-08-698-948-5	Sequence 5, Appl
c 27	52.5	2.2	180	1	US-08-343-281A-20	Sequence 20, Appl
c 28	52.5	2.2	231	3	US-08-905-124-4	Sequence 4, Appl
c 29	52.5	2.2	242	4	US-08-928-213B-65	Sequence 65, Appl
c 30	52.5	2.2	268	1	US-08-105-168B-3	Sequence 3, Appl
c 31	52.5	2.2	268	2	US-08-698-948-3	Sequence 3, Appl
c 32	52.5	2.2	289	4	US-09-065-383-2	Sequence 2, Appl
c 33	52	2.2	256	4	US-08-651-1558-190	Sequence 190, App
c 34	52	2.2	273	4	US-09-134-001C-2418	Sequence 2418, Ap
c 35	51.5	2.2	268	1	US-08-105-168B-1	Sequence 1, Appl
c 36	51.5	2.2	268	2	US-08-105-168B-2	Sequence 2, Appl
c 37	51.5	2.2	268	1	US-08-698-948-1	Sequence 1, Appl
c 38	51.5	2.2	268	2	US-08-698-948-2	Sequence 2, Appl
c 39	51	2.1	186	4	US-09-134-001C-205	Sequence 205, App
c 40	51	2.1	234	2	US-08-611-757-97	Sequence 97, Appl
c 41	51	2.1	234	5	PCT-US95-05980-97	Sequence 97, Appl
c 42	51	2.1	246	2	US-08-673-190A-9	Sequence 9, Appl
c 43	50.5	2.1	100	1	US-08-145-705A-5	Sequence 5, Appl
c 44	50.5	2.1	194	4	US-09-060-756-290	Sequence 290, App
c 45	50.5	2.1	245	4	US-09-397-787-265	Sequence 265, App

ALIGNMENTS

RESULT 1

US-08-651-1558-186
; Sequence 186, Application US/08651155B
; Patent No. 6365401

GENERAL INFORMATION:

APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiethoff Mr., Douglas M.

TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION

TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST

TITLE OF INVENTION: INFECTION

NUMBER OF SEQUENCES: 255

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chrisman, Bynum & Johnson, P.C.

STREET: 1900 Fifteenth Street

CITY: Boulder

STATE: CO

COUNTRY: USA

ZIP: 80302

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/651,155B

FILING DATE: 17-MAY-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Petersen Mr., Steven C.

REGISTRATION NUMBER: 36,238

REFERENCE/DOCKET NUMBER: 17060.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/546-1300

TELEFAX: 303/449-5426

TELEX: ABA1475

INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-651-155B-186

Alignment Scores:
Pred. No.: 1,33e-20 Length: 294
Score: 263.00 Matches: 72
Percent Similarity: 79.5% Conservative: 6
Best Local Similarity: 73.47% Mismatches: 12
Query Match: 10.9% Indels: 10
DB: 4 Gaps: 5

US-09-912-020-325 (1-477) x US-08-651-155B-186 (1-294)

QY 267 IleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhePhe 286
|||||
DB 2 ATCGCGCTGCTGCGCGACCTG-GCCGCGGAAATACCTGGAGAGCGGTATTTC 60
QY 287 AlaAsnAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerProfile 306
|||||
DB 61 GCCATGCGCGCGCGGTAGTGTAGTAACTCGGACGTCACCGTTTCCCTATT 120
QY 307 GluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGluGlu 326
|||||
DB 121 GAGCTGAAACACGAGTCGCGGACG- -GATACCGGCTTCGGCGTTATGACCGAAGAG 176
QY 327 GluLeuLysLeuAlaValAlaAlaAlaArgLysArgGlyGluLysValValMetThrAsn 346
|||||
DB 177 GAGTTGAGACAGCGCTGCCACGCGTAAGTC- -GCCGAGAAAGTGTATGACCAAC 230
QY 347 GlyValPheAspIle- -LeuHis- - -AlaGlyHisValSerThrLeu 360
|||||
DB 231 - -CGCTTCGATATCGACGGCATTATGACGCACTTATCGGACTTA 281

RESULT 2

US-09-060-756-722
Sequence 722, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 722
LENGTH: 280
TYPE: DNA
ORGANISM: Mycobacterium bovis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-722

Alignment Scores:
Pred. No.: 93.4 Length: 280
Score: 65.50 Matches: 36
Percent Similarity: 36.15% Conservative: 11
Best Local Similarity: 27.69% Mismatches: 44

Query Match: 2.74% Indels: 40
DB: 4 Gaps: 6
US-09-912-020-325 (1-477) x US-09-060-756-722 (1-280)
QY 240 LeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluValTyrAspVal 259
|||||
DB 3 CTGGGACCGCCACGGTGCNCNCGCGGTTCGCGACGCTGCTCGCCAGGTG- - - - - 53
QY 260 ThrGlyAlaGlyAspThrValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSer 279
|||||
DB 54 - - - - -TCCGCGCGCGCGTGCATGTGTCCGGATCG- - - - -GCCCTGGCCAGAGC 101
QY 280 LeuGluGluAlaCysPhePheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGly 299
DB 101 - - - - - 101
QY 300 ThrSerThrValSerProIleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGly 319
DB 102 - - - - -CCACCGGTG- - - - -AATCTGGCGCTGGCGACCGCGCGCC 140
QY 320 PheGlyValMetThrCluGluLeuLysLeuAlaValAlaAlaAlaArgLysArg- - - 338
DB 141 GTAGGCTTG- - - - -CGATCGTCACGCGTGGCGTGGCCAGGACGAGATCCCGACGG 191
QY 339 - - -GlyGluLysValValMetThrAsnGlyValPheAspIleLeuHisAlaGlyHisVal 357
DB 192 ATTGGGCG-AGATGCGTGTCTACCATCGGGGTATTGTGACGCGTGCACCGCGGCACGCC 250
QY 358 SerTyrLeuAlaAsnAlaArgLysLeuGly 367
DB 251 GAACTGATCGCGCACGCGGTCAAAGCGGC 280

RESULT 3

US-08-716-942-22/c
Sequence 22, Application US/08716942
Patent No. 5849491
GENERAL INFORMATION:
APPLICANT: Terragen Diversity Inc.
APPLICANT: Radomski, Christopher C. A.
APPLICANT: Seow, Kah Tong
APPLICANT: Warren, R. Anthony J.
APPLICANT: Yap, Wai Ho
TITLE OF INVENTION: METHOD FOR ISOLATING XYLANASE GENE
TITLE OF INVENTION: SEQUENCES FROM SOIL DNA, COMPOSITIONS USEFUL IN SUCH METHOD
TITLE OF INVENTION: COMPOSITIONS OBTAINED THEREBY
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppedahl & Larson
STREET: 1992 Commerce Street, Suite 309
CITY: Yorktown Heights
STATE: NY
COUNTRY: USA
ZIP: 10598-4412
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,942
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/004,157
FILING DATE: 20-Sep-95
ATTORNEY/AGENT INFORMATION:
NAME: Marina T. Larson
REGISTRATION NUMBER: 32,038
REFERENCE/DOCKET NUMBER: TERR.P-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 245-3252
TELEFAX: (914) 962-4330

APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 209
LENGTH: 236
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-209

Alignment Scores:
Pred. No.: 375 Length: 236
Score: 59.00 Matches: 20
Percent Similarity: 38.60% Conservative: 2
Best Local Similarity: 35.09% Mismatches: 19
Query Match: 2.47% Indels: 16
DB: 4 Gaps: 2

US-09-912-020-325 (1-477) x US-09-060-756-209 (1-236)
QY 49 AtgproGlyValAlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArg 68
Db 221 CGCGGGCGGTGGCGGAACATGGCTTCGCCATCGGTGTCTAGGTGGCGAGTCCGG 162
QY 69 LeuValGlyLeuThrGlyIleAspAspAla-----AlaArgAla 81
Db 161 CTAGTTGGGGCGGCGAGCGCGGACTTCGCCGATTATCGCGACTGCGTGAAGCGCGC--- 105
QY 82 LeuSerLysSerLeuAlaAspValAsnValLysCysAspPheValSerVal 98
Db 104 -----GGTGTCAACTGGGAGCACCGTCTCTGATC 78

RESULT 8
US-08-818-112-37/c
Sequence 37, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-37

Alignment Scores:
Pred. No.: 589 Length: 290
Score: 58.50 Matches: 24
Percent Similarity: 42.03% Conservative: 5
Best Local Similarity: 34.78% Mismatches: 23
Query Match: 2.44% Indels: 17
DB: 4 Gaps: 3

US-09-912-020-325 (1-477) x US-08-818-112-37 (1-290)
QY 266 ValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluAlaCysPhe 285
Db 290 GTCACCGCGGTCCACCGCGTCCACCCGAGCGGTGCGCTCTGAGTCCGCC----- 237
QY 286 PheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 236 ---GCCAATNCCGCC-----CTGGCCACCCTTACGCC 207
QY 306 IleGluLeuGluAsnAlaValArgGlyArgAlaAspThrGlyPheGlyValMetThrGlu 325
Db 206 GTTGCCACCAAGCGCGCGTCCGCGGCGTTCCTCCGCC-----ACCGCC 162
QY 326 GluGluLeuLysLeuAlaValAlaAla 334
Db 161 CGCGCGCCCAAGACCGCGTTCGCCGCC 135

RESULT 9
US-08-818-111-37/c
Sequence 37, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

```
;
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-818-111-37
;
Alignment Scores:
Pred. No.: 589
Score: 58.50
Percent Similarity: 42.03%
Best Local Similarity: 34.78%
Mismatch: 2.44%
Indels: 17
Gaps: 3
DB: 4

US-09-912-020-325 (1-477) x US-08-818-111-37 (1-290)
QY 266 ValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhe 285
Db 290 GTCACCGCGTCACCGCGTCACCGCGCGCGTCGCTGAGTGGCGCC----- 237
QY 286 PheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 236 ---GCCAATNCCGCC-----CTGGCCACCCCTTACCGCC 207
QY 306 IleGluLeuGluAsnAlaValArgAlaAlaAspThrGlyPheGlyValMetThrGlu 325
Db 206 GTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
QY 326 GluGluLeuLysLeuAlaValAlaAla 334
Db 161 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 135

RESULT 10
US-09-056-556-37/c
; Sequence 37, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

TREATM

```
;
; TOPOLOGY: linear
; US-09-056-556-37
;
Alignment Scores:
Pred. No.: 589
Score: 58.50
Percent Similarity: 42.03%
Best Local Similarity: 34.78%
Mismatch: 2.44%
Indels: 17
Gaps: 3
DB: 4

US-09-912-020-325 (1-477) x US-09-056-556-37 (1-290)
QY 266 ValIleGlyValLeuAlaAlaThrLeuAlaAlaGlyAsnSerLeuGluGluAlaCysPhe 285
Db 290 GTCACCGCGTCACCGCGTCACCGCGCGCGTCGCTGAGTGGCGCC----- 237
QY 286 PheAlaAsnAlaAlaAlaGlyValValGlyLysLeuGlyThrSerThrValSerPro 305
Db 236 ---GCCAATNCCGCC-----CTGGCCACCCCTTACCGCC 207
QY 306 IleGluLeuGluAsnAlaValArgAlaAlaAspThrGlyPheGlyValMetThrGlu 325
Db 206 GTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
QY 326 GluGluLeuLysLeuAlaValAlaAla 334
Db 161 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 135

RESULT 11
US-09-072-596-37/c
; Sequence 37, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```


Qy 302 Thr 302
: : :
Db 7 AGC 5

Search completed: November 25, 2002, 01:01:12
Job time : 76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: November 24, 2002, 23:21:10 : Search time 76 Seconds
(without alignments)
2377.031 Million cell updates/sec

Title: US-09-912-020-325
Perfect score: 2393
Sequence: 1 MKVTLPEFERGVWVGDM.....FEDGCTNIKKIQDQKG 477

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 335578 seqs, 189365133 residues

Total number of hits satisfying chosen parameters: 345790
Minimum DB seq length: 0
Maximum DB seq length: 300
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPFO.spool/US09912020/runat_20112002_150301_11146/app_query.fasta_1.647
-DB=Published-Applications_NA -QFMT=fastap -SUFFIX=p2nszlm300.rnpb
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1
-MATRIX=blosum62 -TRANS=human40 -cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=300
-USER=US09912020 -CGN_1_1_47@runat_20112002_150301_11146 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Match Length DB ID	Description
c 1	120	5.0	272 10 US-09-741-669-89 Sequence 89, Appl
c 2	112	4.7	292 10 US-09-917-800A-1264 Sequence 1264, Ap
c 3	107.5	4.5	285 10 US-09-294-093B-5061 Sequence 5061, Ap

4	92	3.8	263	10	US-09-878-574-10531	Sequence 10531, A
5	82.5	3.4	265	10	US-09-878-574-8726	Sequence 8726, Ap
c 6	70	2.9	245	10	US-09-741-669-90	Sequence 90, Appl
7	69	2.9	272	10	US-09-923-876-159	Sequence 159, App
8	67.5	2.8	281	10	US-09-294-093B-3810	Sequence 3810, Ap
9	65	2.7	287	10	US-09-878-574-2532	Sequence 2532, Ap
10	64	2.7	272	9	US-10-046-935-1960	Sequence 1960, Ap
c 11	64	2.7	298	10	US-09-294-093B-5568	Sequence 5568, Ap
c 12	63.5	2.7	267	10	US-09-864-761-27168	Sequence 27168, A
13	63	2.6	268	10	US-09-923-876-729	Sequence 729, App
14	61	2.5	176	9	US-09-954-531-285	Sequence 285, App
15	60.5	2.5	234	10	US-09-974-300-7643	Sequence 7643, Ap
16	60.5	2.5	254	10	US-09-878-574-6039	Sequence 6039, Ap
17	60.5	2.5	256	10	US-09-923-876-2372	Sequence 2372, Ap
c 18	60.5	2.5	291	10	US-09-134-333-1	Sequence 1, Appl
19	59.5	2.5	248	10	US-09-974-300-8309	Sequence 8309, Ap
20	59.5	2.5	294	10	US-09-815-242-7907	Sequence 7907, Ap
c 21	59	2.5	117	10	US-09-864-761-28313	Sequence 28313, A
22	59	2.5	238	10	US-09-864-761-24079	Sequence 24079, A
c 23	59	2.5	296	9	US-09-933-797-461	Sequence 461, App
24	58	2.4	261	10	US-09-864-761-26632	Sequence 26632, A
25	58	2.4	261	10	US-09-793-306-159	Sequence 159, App
26	58	2.4	276	10	US-09-878-574-10955	Sequence 10955, A
27	58	2.4	297	10	US-09-920-300A-220	Sequence 220, App
28	58	2.4	297	12	US-10-033-528-220	Sequence 220, App
c 29	57.5	2.4	211	10	US-09-867-701-6838	Sequence 6838, Ap
c 30	57.5	2.4	232	10	US-09-878-574-6908	Sequence 6908, Ap
c 31	57.5	2.4	271	10	US-09-878-574-7082	Sequence 7082, Ap
c 32	57.5	2.4	277	10	US-09-960-352-12295	Sequence 12295, A
33	57	2.4	148	10	US-09-864-761-17837	Sequence 17837, A
34	57	2.4	228	10	US-09-864-761-19281	Sequence 19281, A
35	57	2.4	273	10	US-09-878-574-14311	Sequence 14311, A
36	57	2.4	273	10	US-09-960-352-1754	Sequence 1754, Ap
37	57	2.4	275	10	US-09-294-093B-2895	Sequence 2895, Ap
38	57	2.4	278	10	US-09-923-876-5571	Sequence 5571, Ap
c 39	56.5	2.4	249	10	US-09-864-761-26634	Sequence 26634, A
40	56.5	2.4	265	9	US-10-046-935-1087	Sequence 1087, Ap
c 41	56.5	2.4	267	10	US-09-974-300-6945	Sequence 6945, Ap
c 42	56.5	2.4	271	10	US-09-294-093B-4387	Sequence 4387, Ap
c 43	56.5	2.4	287	10	US-09-878-574-13727	Sequence 13727, A
c 44	56.5	2.4	288	9	US-10-046-935-598	Sequence 598, App
c 45	56.5	2.4	288	9	US-10-046-935-1291	Sequence 1291, Ap

ALIGNMENTS

RESULT 1
US-09-741-669-89/c
; Sequence 89, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for proliferation of E. coli
; FILE REFERENCE: ELITRA 009A
; CURRENT APPLICATION NUMBER: US/09/741.669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-669-89
Alignment Scores:
Pred. No.: 3.66e-05 Length: 272
Score: 120.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00%
Query Match:
DB:
    Mismatches: 0
    Indels: 0
    Gaps: 0
```

US-09-912-020-325 (1-477) x US-09-741-669-89 (1-272)

Qy	455	valLeuAsnPheGluAspGlyCysSerThrThrAsnIleIleIleLysLysLysIleGlnGlnAsp	474
Db	271	GTGCTCAACTTTGAAGACGGTTGCTCGACGACCACATCATCAAGAAGATGCCAACAGGAT	212
Qy	475	LysLysGly	477
Db	211	AAAAAAGCC	203

RESULT 2

```

US-09-917-800A-1264/c
: Sequence 1264, Application US/09917800A
: Patent No. US20020119462A1
: GENERAL INFORMATION:
: APPLICANT: Mendrick, Donna
: APPLICANT: Porter, Mark
: APPLICANT: Johnson, Kory
: APPLICANT: Castie, Arthur
: APPLICANT: Elashoff, Michael
: APPLICANT: Gene Logic, Inc.
: TITLE OF INVENTION: Molecular
: FILE REFERENCE: 44921-5038-US
: CURRENT APPLICATION NUMBER: US/09/917,800A
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/222,040
: PRIOR FILING DATE: 2000-07-31
: PRIOR APPLICATION NUMBER: US 60/222,880
: PRIOR FILING DATE: 2000-11-02
: PRIOR APPLICATION NUMBER: US 60/290,029
: PRIOR FILING DATE: 2001-05-11
: PRIOR APPLICATION NUMBER: US 60/290,645
: PRIOR FILING DATE: 2001-05-15
: PRIOR APPLICATION NUMBER: US 60/292,336
: PRIOR FILING DATE: 2001-05-22
: PRIOR APPLICATION NUMBER: US 60/295,798
: PRIOR FILING DATE: 2001-06-06
: PRIOR APPLICATION NUMBER: US 60/297,457
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/298,884
: PRIOR FILING DATE: 2001-06-19
: PRIOR APPLICATION NUMBER: US 60/303,459
: PRIOR FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 1740
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1264

```

OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI236089
US-09-917-800A-1264

US-09-912-020-325 (1-477) x US-09-917-800A-1264 (1-292)

Qy	237	MetSerLeuLeuGlnProGlyLysAlaProLeuHisMetProThrGlnAlaGlnGluVal	256
		
Db	278	GTGACATGTCACAGGCAGAACCTGTTCCAAAGCATTCCACAGACGTCAGGCT	219
		
Qy	257	TyrAspValThrGlyAlaGlyAspThrValIleGlyValLeuAlaIaThrLeuAla---	275

Db	218	GTGGACACACCGGTGCTGGT	GACAGTTTTTGTGGGAGCGCTTGCC	TTCTACCTACCTGGCTTAC	159
Qy	276	---AlaGlyAsnSerLeuGluGluL	uLaCysPhePheAlaAsnAlaAla	AlaGlyValVal	294
Db	158	TACCCAAAGTCGTCTGTGG	AAGAAATGCTCAAGAGATCTAA	TTCCATCGCTGGGTACG	99
Qy	295	ValGlyLysLeuGlyThrSerThr	ValSerPro	305	
Db	98	GTCAGGCCACAGAAACACAG	TCCCTCTTATCCA	66	

RESULT 3

```

US-09-294-093B-5061
; Sequence 5061, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294, 093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5061
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700355719H1
; NAME/KEY: unsure
; LOCATION: 83, 257, 263, 278
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-5061

```

Alignment Scores:	
Pred. No.:	0.000967
Score:	107.50
Percent Similarity:	43.82%
Best Local Similarity:	29.21%
Query Match:	4.49%
DB:	10
Length:	285
Matches:	26
Conservative:	13
Mismatches:	27
Indels:	23
Gaps:	1

US-09-912-020-325 (1-477) x US-09-294-093B-5061 (1-285)

Qy	316	AlaAspThrGlyPheGlyValMetThrGluGlu	-----	327
Db	1	GCTCACACTGATTAACTGTTGCTGAGAAGCTCGAGAGAAATCAATGATGAA	60	
Qy	328	-----	-----	332
		-----LeuLysLeuAlaVal		
Db	61	GTGAAAGCTCAGCTATCTCATNCCCTTCACACTTCGCGGGATCAGCAGTTTTCAAAT	120	
Qy	333	AlaAlaArgLysArgGlyLysValValMetThrAsnGlyValPheAspLeu	352	
Db	121	GGCGAGCGCTCCTCGCCAGGTGCTGTGTGTATGTAGATGGCACATTTGATCTTTTC	180	
Qy	353	HisAlaGlyHisValSerTyrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuVal	372	
Db	181	CACGCTGCCCATGTTGAGTTCCTCAGGAGTGCCAGACACTTGGTGACCTCTCTTTGTGG	240	
Qy	373	AlaValAsnSerAspAlaSerThrLys	381	
Db	241	GTATCTATGGACGACGNATCGANTCAG	267	

RESULT 4

US-09-878-574-10531
; Sequence 10531, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:

```

; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10531
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700966779H1
US-09-878-574-10531

Alignment Scores:
Pred. No.: 0.046 Length: 263
Score: 92.00 Matches: 22
Percent Similarity: 58.93% Conservative: 11
Best Local Similarity: 39.29% Mismatches: 23
Query Match: 3.84% Indels: 1
DB: 10 Gaps: 0

US-09-912-020-325 (1-477) x US-09-878-574-10531 (1-263)

Qy 355 GlyHisValSerThrLeuAlaAsnAlaArgLysLeuGlyAspArgLeuIleValAlaVal 374
|||||
2 GCCTGTTGAGTACTTAAAGAGGCTAGGAGCTTGTGACTTCTTCTAGTTGGTATC 61

Qy 375 AsnSerAspAlaSerThrLysArgLeuLysGlyAspSerArgProValAsnProLeuGlu 394
|||||
62 CACTCGATGAGACAGTACGAGTACGATGAGGAAATCACTATCCAAATTATGATCTGCAT 121

Qy 395 GlnArgMetIleValLeuGlyAlaLeuGluAlaValAlaAspTrpValVal 410
|||||
122 GAGCGTACCTAGTGTGTTA-GCTTGGCGTTATGTTGATGAAGTTATT 168

RESULT 5
US-09-878-574-8726
; Sequence 8726, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 8726
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701101576H1
US-09-878-574-8726

Alignment Scores:
Pred. No.: 0.533 Length: 265
Score: 82.50 Matches: 20
Percent Similarity: 61.22% Conservative: 10
Best Local Similarity: 40.82% Mismatches: 18
Query Match: 3.45% Indels: 1
DB: 10 Gaps: 1

US-09-912-020-325 (1-477) x US-09-878-574-8726 (1-265)

```

```

Qy 50 ProGlyGlyAla---AlaAsnValAlaMetAsnIleAlaSerLeuGlyAlaAsnAlaArg 68
|||||
Db 51 CCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 110
|||||
Qy 69 LeuValGlyLeuThrGlyIleAspAlaAlaArgAlaLeuSerLysSerLeuAlaAsp 88
|||||
Db 111 TTCGTCGGGAAAGTCGGGCGAGGAGTTCGGGCGACATGCTGGCCCGAATCCTGAAGGAG 170
|||||
Qy 89 ValAsnValLysCysAspPheValSer 97
|||||
Db 171 AACGAGCTGCGATCCGACGGGATCAAC 197
|||||

RESULT 6
US-09-741-669-90/c
; Sequence 90, Application US/09741669
; Patent No. US2002022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 90
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-669-90

Alignment Scores:
Pred. No.: 11.8 Length: 245
Score: 70.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.93% Indels: 0
DB: 10 Gaps: 0

US-09-912-020-325 (1-477) x US-09-741-669-90 (1-245)

Qy 464 ThrThrAsnIleIleLysLysIleGlnGlnAspLysLysGly 477
|||||
Db 244 ACGACCAACATCATCAAGAAGATCCACAGGATAAAAAAGGC 203
|||||

RESULT 7
US-09-923-876-159
; Sequence 159, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalquid, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 159
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature

```


Search completed: November 25, 2002, 01:02:42
Job time : 78 secs

